Bilateral Bargaining with a Biased Intermediary *

Yusuke Yamaguchi†

August 7, 2025

Abstract

Bilateral bargaining is often facilitated by an intermediary. In many settings, however, the intermediary is biased—sharing interests with one of the negotiating parties—and lacks both commitment and enforcement power. This paper examines how such a biased intermediary affects bargaining outcomes. I consider a stylized bilateral trade framework and compare two bargaining games: a seller-offer bargaining game, in which the seller proposes a price, and a mediated bargaining game, in which the intermediary proposes a price and traders pay her commissions. By focusing on the set of communication equilibria in both games, I characterize the bounds on expected social surplus achievable in equilibrium when the players are allowed to engage in general preplay and intraplay communication. I show that when the commission cost is sufficiently small, the mediated bargaining game can yield a higher expected social surplus than the seller-offer bargaining game in the second-best scenario. This result provides a rationale for the widespread use of intermediaries in bargaining, even when their bias is common knowledge.

Keywords: Bargaining, Intermediary, Bias, Communication equilibrium.

JEL Classification: C78, D82.

*This paper won the Kanematsu Prize in 2024 from the Research Institute for Economics and Business Administration, Kobe University. I am grateful to Takuro Yamashita, Johannes Hörner, and David Martimort for their invaluable guidance and support. I also thank Milo Bianchi, Doh-Shin Jeon, Shintaro Miura, Yasuyuki Miyahara, François Salanié, Anna Sanktjohanser, Sara Shahanaghi, Alex Smolin, and audiences at CPW, AMES China 2022, ESAM 2022, CMiD 2022, AMES 2022 Tokyo, the 2024 JEA Autumn Meeting, and the Frontiers in Behavioral and Experimental Economics Workshop, as well as the anonymous referees of the Kanematsu Prize, for their helpful comments. Part of this research was conducted while I was visiting the Osaka School of International Public Policy, The University of Osaka, whose hospitality I gratefully acknowledge. Financial support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 714693) and JSPS KAKENHI (JP23K01311) is also gratefully acknowledged. The usual disclaimer applies.

†The University of Osaka, Osaka, Japan; and Toulouse School of Economics, Toulouse, France. yusuke.yamaguchi36@gmail.com

1 Introduction

Bilateral bargaining is often facilitated by an *intermediary* who communicates with both negotiating parties and offers them a resolution. By obfuscating the private information that direct communication would otherwise reveal, such mediation is believed to reduce strategic incentives and promote efficient outcomes.

However, in many real-world situations, the intermediary is not benevolent but instead shares interests with one of the negotiating parties. She may behave opportunistically to secure outcomes aligned with her own objectives, potentially harming the efficiency of the bargaining outcome if these objectives conflict with efficiency.[1] Moreover, she is not omnipotent: she cannot commit to or enforce her decisions. This lack of commitment and enforcement power further limits the effectiveness of mediation.

For example, transactions of real estate, artworks, or used cars often involve an agent who facilitates trade by communicating with a seller and a buyer and offering them a price. As she earns a commission proportional to the sale price, she has the incentive to realize a higher sale price and is considered *biased* toward the seller.[2] Typically, the agent cannot commit to the price in advance, and trade occurs if and only if both the seller and the buyer accept the offer.

The purpose of this paper is to study how having a *biased intermediary who lacks both commitment and enforcement power* affects the bargaining outcome. To this end, I consider a stylized bilateral trade setting and compare two bargaining games: one in which the seller makes an offer and one in which a biased intermediary does so. In both games, the seller owns a good and seeks to trade it with the buyer. Each trader's valuation of the good is binary and independently distributed, and its realization is privately known to the trader. There is always a gain from trade except for the case where the seller is high type and the buyer is low type. In the *seller-offer bargaining game*, payoffs are given by standard linear payoffs: if trade occurs, each trader obtains the difference between the price and their valuation; if no trade occurs, both obtain zero. In the *mediated bargaining game*, if trade occurs, each trader also pays a fixed-rate

[1] Throughout the paper, I use feminine pronouns for the intermediary and masculine pronouns for the seller and the buyer, who will appear shortly.

[2] For example, in the U.S. real estate industry, it has been standard for the seller to pay about 5%. See Section 5 for related discussion.

commission proportional to the price, and the intermediary's payoff is given by the sum of these commissions. If no trade occurs, all players obtain zero.

To characterize the bounds on expected social surplus under general preplay and intraplay communication, both games incorporate *communication devices*. A communication device is characterized by a *mediation plan*, which outputs action recommendations based on reported types. The mediated bargaining game proceeds as follows: First, the seller and the buyer observe the realizations of their type. Second, they privately report their type to the communication device. Third, the communication device privately recommends a price to the intermediary and a minimum (resp. maximum) acceptable price to the seller (resp. the buyer). Fourth, the intermediary offers a price to the traders. Finally, the traders simultaneously respond to the offer by either "Accept" or "Reject." If both traders accept the offer, trade occurs at that price; if at least one of them rejects it, no trade occurs.[3]

An incentive-compatible mediation plan is called a *communication equilibrium* (*CE*). Forges (1986) shows that the set of CEs characterizes the set of outcomes achievable through some form of preplay or intraplay communication. However, certain CEs are not robust to standard trembling-hand-type perturbations. To address this, I propose a refinement called *acceptable* CE, which requires that the communication device always recommends the "true" minimum and maximum acceptable prices given the reported types. In other words, it recommends accepting an offer if and only if it guarantees a nonnegative payoff.

The main result of the paper is that if the commission cost is sufficiently small, in the second-best (SB) scenario, the mediated bargaining game can yield a higher ex ante expected social surplus than the seller-offer bargaining game. I establish this by deriving necessary conditions for acceptable CEs in both games. Roughly, these conditions imply that the player making the offer always proposes one of the buyer's maximum acceptable prices.[4] Intuitively, under acceptable CE, traders accept any mutually acceptable price—a price that yields nonnegative payoffs to both traders—so the player with bargaining power has the incentive to offer as high a price as possible. Hence, whenever there is a gain from trade, there are only two possible prices (low or high) that

[3]In the seller-offer bargaining game, a mediation plan recommends a price to the seller and a maximum acceptable price to the buyer. After the seller offers a price, only the buyer responds to it, and trade occurs if and only if he accepts.

[4]Since the buyer has two possible types, there are two such prices.

can be offered in equilibrium. This implies that the players' IC constraints can be written as linear inequalities in the total probabilities that the high price is offered for each type profile. As such, the SB outcome can be characterized by solving the corresponding linear program.

In the seller-offer bargaining game, the SB outcome entails the seller always offering the high price whenever there is a gain from trade. In the mediated bargaining game, the corresponding linear program implies a threshold for the ratio of the buyer's total payment to the seller's revenue, beyond which the SB outcome shifts regime. This ratio is increasing in both commission rates and thus indirectly reflects the commission cost, or equivalently, the intermediary's commission gain. If the ratio exceeds the threshold, the commission gain is too large to ignore, and the intermediary cannot be incentivized to offer the low price. As a result, she always offers the high price whenever there is a gain from trade—mirroring the seller's behavior in the SB outcome of the seller-offer bargaining game. Consequently, the SB levels of ex ante expected social surplus coincide under the two games. In this case, mixing the low and high prices would lead to a higher expected social surplus, yet the intermediary is not willing to do so.

However, if the ratio is below the threshold, the intermediary can be incentivized to offer the low price to the high-type pair, resulting in strictly higher expected social surplus. This is possible because she does not know the seller's type and believes that it may be accepted with positive probability. In contrast, the seller knows his own type and would not offer a price that would yield him a negative payoff, which is a key distinction between the seller and the intermediary. Therefore, when the commission cost is sufficiently small, the mediated bargaining game can strictly outperform the seller-offer bargaining game in terms of expected social surplus. This result provides a rationale for the widespread use of intermediaries in bargaining, even when their bias is common knowledge.

Finally, it is worth noting that the mediated bargaining game studied here represents a minimal departure from the seller-offer bargaining game in that the intermediary is *weak*. That is, she shares interests with the seller, possesses exactly the same instrument—namely, the ability to offer a price rather than a binding contract or a mechanism—and has no private information or expertise. Yet, the main result demonstrates that even such a weak intermediary can improve the efficiency of the bargaining outcome.

1.1 Related literature

Three features of the intermediary—bias, lack of commitment, and lack of enforcement—distinguish this paper from existing work. Intermediaries, broadly defined, have been studied in the mechanism design literature.[5] Classic papers such as Myerson and Satterthwaite (1983) can be viewed as studying the bargaining with an unbiased intermediary with commitment and enforcement power: the principal maximizes expected surplus, can commit to a mechanism, and the agents' acceptance is not necessary once they have agreed to participate.[6] While studying a *weak* intermediary offers a natural and minimal departure from bilateral bargaining to examine the effect of biased mediation, much of the literature uses Myerson and Satterthwaite (1983) as a starting point and explores various departures from it. In doing so, the three features above have typically been studied separately. For example, papers on optimal mechanism design by the seller, such as Myerson (1981), can be viewed as examining an extreme case of a biased intermediary whose preferences are completely aligned with the seller's. Similarly, Myerson (1982) analyzes mechanism design problems involving both hidden information and hidden actions, effectively modeling a principal without enforcement power. Regarding the commitment assumption, there is a growing literature on mechanism design with limited commitment (see Bester and Strausz, 2001, 2000, 2007; Doval and Skreta, 2022; Lomys and Yamashita, 2022). Relatedly, Eilat and Pauzner (2021) study bilateral trade with a benevolent intermediary who lacks commitment power.[7] The novelty of this paper is thus to consider all three features simultaneously.[8] Combined with the lack of commitment, the intermediary's bias creates incentives to offer as high a price as possible, which is the driving force of the results.

A biased intermediary has also been studied in the context of international relations (see Kydd, 2003, 2006). Kydd (2003) shows that an intermediary can reduce the probability of conflict only if he is biased, implying that an unbiased intermediary is ineffective.[9]

[5] An intermediary in this literature is typically called a *principal*.

[6] In this sense, the lack of enforcement power means that individual rationality (IR) constraint must be satisfied ex post rather than ex ante.

[7] They directly analyze the game between the traders and the intermediary in which the intermediary offers a mechanism. As the intermediary in my model offers a price rather than a contract or mechanism, Eilat and Pauzner (2021) allow more flexibility in the intermediary's actions.

[8] Some papers combine two of these features. For example, Lomys and Yamashita (2022) and Doval and Skreta (2024) analyze seller-optimal mechanisms under limited commitment.

[9] In contrast, an unbiased intermediary always outperforms a biased intermediary in this paper (see Proposition 1 and footnote 25).

This paper also differs from some previous work in that the intermediary is modeled as an active player of the game, who has a preference and actions to choose from. This clarifies how the intermediary's bias affects her facilitative role—namely, facilitating agreement by obfuscating the private information that direct communication would otherwise reveal.[10] See Čopič and Ponsatí (2008); Fanning (2021, 2023); Jarque, Ponsatí and Sákovics (2003) for studies focusing on this facilitative role, and Ganguly and Ray (2023); Goltsman, Hörner, Pavlov and Squintani (2009); Hörner, Morelli and Squintani (2015) for other work on mediation.

Methodologically, this paper builds on the work of Forges (1986), Myerson (1986b), and Sugaya and Wolitzky (2021), who study multistage games with communication and establish the *communication revelation principle* for various equilibrium concepts. In the context of this paper, the communication revelation principle implies that the set of CEs characterizes the set of outcomes achievable through some form of preplay or intraplay communication between the players. In particular, it is without loss of generality to focus on the canonical communication devices described above, which output action recommendations based on reported types.

The remainder of the paper is organized as follows. Section 2 describes the model and defines CE and acceptable CE. Section 3 presents some preliminary results. Section 4 provides the main result of the paper. Section 5 discusses alternative payoff specifications and concludes the paper. Some proofs are provided in Appendix.

2 Model

Environment. I study two bargaining environments: one that involves only a seller and a buyer, and another that also involves an intermediary. In both environments, the seller owns a good and seeks to trade it with the buyer. Each trader's valuation of the good is binary and independently distributed, and its realization is privately known to the trader. Specifically, the seller's valuation is high (s_H) with probability $\pi_S \in (0, 1)$, and low (s_L) with probability $1 - \pi_S$. Similarly, the buyer's valuation is high (b_H) with probability $\pi_B \in (0, 1)$, and low (b_L) with probability $1 - \pi_B$. Let $\Theta_S = \{s_L, s_H\}$ and $\Theta_B = \{b_L, b_H\}$. There is a gain from trade for all type profiles except

[10] Gottardi and Mezzetti (2024) study not only this facilitative role but also the evaluative role; that is, providing guidance on an appropriate resolution based on expertise.

(s_H, b_L); that is, $0 < s_L < b_L < s_H < b_H$.

Preferences. In the seller-offer bargaining game (defined in Section 2.1), each player's payoff is given by the difference between their valuation and the trading price: if the traders' types are (s, b) and they trade at a price p, the seller obtains $p - s$ and the buyer obtains $b - p$; if no trade occurs, both obtain zero.

In the mediated bargaining game (defined in Section 2.2), both the seller and the buyer pay commissions proportional to the trading price. Specifically, if the traders' types are (s, b) and they trade at a price p, the seller obtains $(1 - \delta_S)p - s$, the buyer obtains $b - (1 + \delta_B)p$, and the intermediary obtains $(\delta_S + \delta_B)p$, where $\delta_S, \delta_B \in (0, 1)$ are fixed commission rates. If no trade occurs, all players obtain zero. Let $h = \frac{1+\delta_B}{1-\delta_S}$. This represents the ratio of the buyer's total payment $(1 + \delta_B)p$ to the seller's revenue $(1 - \delta_S)p$. As h increases with the commission rates, it captures the cost of employing the intermediary. Let $\bar{h} = \min\left\{\frac{b_L}{s_L}, \frac{b_H}{s_H}\right\}$. I assume $h \leq \bar{h}$, so that a *mutually acceptable* price exists if and only if there is a gain from trade.[11] Since the intermediary's payoff is increasing in the price p, she shares interests with the seller and is thus considered *biased*.

2.1 Seller-offer bargaining game

As a benchmark, I first consider the *seller-offer bargaining game*, in which the seller gives a take-it-or-leave-it price offer to the buyer. The game incorporates a *communication device* that recommends actions to both players based on the reported types. A *pure mediation plan* is a pair of functions (q, r^{SO}), where $q\colon \Theta_S \times \Theta_B \to \mathbb{R}_+$ is a *price recommendation* to the seller, and $r^{\text{SO}}\colon \Theta_S \times \Theta_B \to \mathbb{R}_+$ is a *response recommendation* to the buyer.[12] For instance, given reported types $(\tilde{s}, \tilde{b}) \in \Theta_S \times \Theta_B$, the communication device recommends that the seller offer a price $q(\tilde{s}, \tilde{b})$ and that the buyer accept the offer if and only if the price does not exceed $r^{\text{SO}}(\tilde{s}, \tilde{b})$; that is, r^{SO} recommends a maximum acceptable price. Let Q and R^{SO} denote the sets of all price and response recommendations, respectively. A communication device is characterized by a

[11] A price p is *mutually acceptable* for a type-s seller and a type-b buyer if both players obtain nonnegative payoffs from trading at the price p. In the mediated bargaining game, this is equivalent to $p \in \left[\frac{s}{1-\delta_S}, \frac{b}{1+\delta_B}\right]$. Provided $h \leq \bar{h}$, this interval is nonempty for all type profiles except (s_H, b_L). See also footnote 18.

[12] Throughout the paper, the superscripts "SO" and "MB" refer to the seller-offer and the mediated bargaining games, respectively.

mediation plan $\mu^{\mathrm{SO}} \in \Delta(Q \times R^{\mathrm{SO}})$—a probability distribution over the set of pure mediation plans—which is common knowledge among the players.

The game proceeds as follows:

1. The seller and the buyer privately observe their type. Let (s, b) denote the realized type profile.

2. They privately report their type $\tilde{s} \in \Theta_S$ and $\tilde{b} \in \Theta_B$ to the communication device.

3. A pure mediation plan $(q, r^{\mathrm{SO}}) \in Q \times R^{\mathrm{SO}}$ is drawn with probability $\mu^{\mathrm{SO}}(q, r^{\mathrm{SO}})$, but the players do not observe which one is drawn.

4. The communication device privately recommends a price $q(\tilde{s}, \tilde{b})$ to the seller and a maximum acceptable price $r^{\mathrm{SO}}(\tilde{s}, \tilde{b})$ to the buyer.

5. The seller offers a price $\tilde{p} \in \mathbb{R}_+$ to the buyer.

6. The buyer accepts or rejects the offer. If accepted, trade occurs at the price $\tilde{p}$: the seller obtains $\tilde{p} - s$, and the buyer obtains $b - \tilde{p}$. If rejected, no trade occurs and both players obtain a payoff of zero.

2.2 The mediated bargaining game

In the *mediated bargaining game*, the intermediary gives a take-it-or-leave-it price offer to the traders, who then simultaneously decide whether to accept it. This game is also equipped with a communication device, though its function differs slightly from that in the seller-offer game. It recommends a price to the intermediary, a minimum acceptable price to the seller, and a maximum acceptable price to the buyer. Formally, the response recommendation now maps into $\mathbb{R}^2_+$; that is, $r^{\mathrm{MB}}(s, b) = \left(r^{\mathrm{MB}}_S(s, b), r^{\mathrm{MB}}_B(s, b)\right)$, where r^{MB}_S and r^{MB}_B are recommendations to the seller and the buyer, respectively. Let R^{MB} denote the set of all such functions, and consider a mediation plan $\mu^{\mathrm{MB}} \in \Delta(Q \times R^{\mathrm{MB}})$.

The game proceeds as follows:

1. The seller and the buyer privately observe their type. Let (s, b) denote the realized type profile.

2. They privately report their type $\tilde{s} \in \Theta_S$ and $\tilde{b} \in \Theta_B$ to the communication device.

3. A pure mediation plan $(q, r^{\text{MB}}) \in Q \times R^{\text{MB}}$ is drawn with probability $\mu^{\text{MB}}(q, r^{\text{MB}})$, but the players do not observe which one is drawn.

4. The communication device privately recommends a price $q(\tilde{s}, \tilde{b})$ to the intermediary, a minimum acceptable price $r_S^{\text{MB}}(\tilde{s}, \tilde{b})$ to the seller, and a maximum acceptable price $r_B^{\text{MB}}(\tilde{s}, \tilde{b})$ to the buyer.

5. The intermediary offers a price $\tilde{p} \in \mathbb{R}_+$.

6. The seller and the buyer simultaneously decide whether to accept the offer. If both accept, trade occurs at the price $\tilde{p}$: the seller obtains $(1 - \delta_S)\tilde{p} - s$, the buyer obtains $b - (1 + \delta_B)\tilde{p}$, and the intermediary obtains $(\delta_S + \delta_B)\tilde{p}$. If either party rejects, no trade occurs and all players obtain a payoff of zero.

2.3 Equilibrium concept

A mediation plan is a *communication equilibrium* (CE) if no player can, ex ante, expect to gain by misreporting their type or disobeying recommendations. The *communication revelation principle* implies that the set of CEs characterizes the set of outcomes achievable through some form of preplay or intraplay communication between the players. In other words, it is without loss of generality to assume a *canonical* communication device described above, which takes the players' types as inputs and recommends an action to each player.[13] In this paper, I introduce the notion of *acceptable communication equilibrium* (acceptable CE) as a refinement of CE. A CE is said to be *acceptable* if its response recommendation prescribes the prices at which the traders break even, given their reported types. Note that players who truthfully report their type find it optimal to follow such recommendations, as doing so guarantees nonnegative expected payoffs.

This refinement is motivated by the standard trembling-hand argument. To see this, consider a pure mediation plan (q, r^{SO}) such that $q(s, b) = p < b$ and $r^{\text{SO}}(s, b) = p$. In this case, the players have no incentive to deviate. In particular, the seller does not deviate because any price $\tilde{p} \in (p, b]$ would be rejected. However, if the seller mistakenly offers some price $\tilde{p} \in (p, b)$,

[13] See Remark 1 for the applicability of the communication revelation principle to the present framework.

the buyer is better off accepting it, implying that obedience to (q, r^{SO}) is not robust to such trembles.[14] By contrast, an acceptable CE is robust to such perturbations.[15]

2.3.1 Acceptable CE in the seller-offer bargaining game

In the seller-offer bargaining game, the above-described response recommendation is captured by $r^{\text{SO}*}$, defined as $r^{\text{SO}*}(s,b) = b$ for all $(s,b) \in \Theta_S \times \Theta_B$.

Let $U_i^{\text{SO}}(q)$ denote the ex ante expected payoff of trader $i \in \{S, B\}$ under pure mediation plan $(q, r^{\text{SO}*})$ when all players are honest and obedient:

$$U_S^{\text{SO}}(q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[q(s,b) - s] \cdot \mathbf{1}_{\{q(s,b)\le b\}},$$
$$U_B^{\text{SO}}(q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[b - q(s,b)] \cdot \mathbf{1}_{\{q(s,b)\le b\}},$$

where $\Pr(s,b)$ is the prior probability that the traders' types are (s,b) and $\mathbf{1}_{\{\cdot\}}$ is the indicator function.

A player may manipulate a mediation plan either by misreporting their type or by disobeying the recommendation. The seller's manipulation is represented by $\sigma_S^{\text{SO}} = \left(\sigma_{S1}^{\text{SO}}, \sigma_{S2}^{\text{SO}}\right)$, where $\sigma_{S1}^{\text{SO}} \colon \Theta_S \to \Theta_S$ is a manipulation in report and $\sigma_{S2}^{\text{SO}} \colon \Theta_S \times \mathbb{R}_+ \to \mathbb{R}_+$ is a manipulation in offer. Let Σ_S^{SO} denote the set of the seller's manipulations. If the traders' types are (s,b) and the seller manipulates $(q, r^{\text{SO}*})$ using $\sigma_S^{\text{SO}} \in \Sigma_S^{\text{SO}}$ while the buyer is honest and obedient, she reports $\sigma_{S1}^{\text{SO}}(s)$, is recommended $q\left(\sigma_{S1}^{\text{SO}}(s), b\right)$, and offers $\tilde{p}(s,b) \equiv \sigma_{S2}^{\text{SO}}\left(s, q\left(\sigma_{S1}^{\text{SO}}(s), b\right)\right)$. The seller's ex ante expected payoff from such manipulation is thus

$$U_S^{\text{SO}}\left(q \circ \sigma_S^{\text{SO}}\right) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[\tilde{p}(s,b) - s] \cdot \mathbf{1}_{\{\tilde{p}(s,b)\le b\}}.$$

The buyer's manipulation is represented by $\sigma_B^{\text{SO}} = \left(\sigma_{B1}^{\text{SO}}, \sigma_{B2}^{\text{SO}}\right)$, where $\sigma_{B1}^{\text{SO}} \colon \Theta_B \to \Theta_B$ is a manipulation in report and $\sigma_{B2}^{\text{SO}} \colon \Theta_B \times \mathbb{R}_+ \to \mathbb{R}_+$ is a manipulation in response. Let Σ_B^{SO} denote the set of the buyer's manipulations. If the traders' types are (s,b) and the buyer

[14] A similar argument can be made for the mediated bargaining game.

[15] This resonates with the notion of *acceptable correlated equilibrium* introduced by Myerson (1986a), hence the name.

manipulates $(q, r^{\text{SO}*})$ using $\sigma_B^{\text{SO}} \in \Sigma_B^{\text{SO}}$ while the seller is honest and obedient, the buyer reports $\tilde{b}(b) \equiv \sigma_{B1}^{\text{SO}}(b)$, is recommended $r^{\text{SO}*}(s, \tilde{b}(b)) = \tilde{b}(b)$, and accepts an offer if and only if the price is smaller than or equal to $p(b) \equiv \sigma_{B2}^{\text{SO}}(b, \tilde{b}(b))$.[16] The buyer's ex ante expected payoff from such manipulation is thus

$$U_B^{\text{SO}}(q \circ \sigma_B^{\text{SO}}) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)\left[b - q\left(s, \tilde{b}(b)\right)\right] \cdot \mathbf{1}_{\{q(s,\tilde{b}(b))\le p(b)\}}.$$

A mediation plan $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$ is an acceptable CE if no player has a profitable manipulation.[17]

Definition 1. In the seller-offer bargaining game, a mediation plan $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$ is an *acceptable communication equilibrium* if, for all $i \in \{S, B\}$ and all $\sigma_i^{\text{SO}} \in \Sigma_i^{\text{SO}}$,

$$\sum_{q\in Q} \mu^{\text{SO}}(q) U_i^{\text{SO}}(q) \ge \sum_{q\in Q} \mu^{\text{SO}}(q) U_i^{\text{SO}}(q \circ \sigma_i^{\text{SO}}). \tag{2.1}$$

2.3.2 Acceptable CE in the mediated bargaining game

In the mediated bargaining game, a type-s seller breaks even at the price $p_S(s) \equiv \frac{s}{1-\delta_S}$ and a type-b buyer does so at $p_B(b) \equiv \frac{b}{1+\delta_B}$.[18] Hence, the response recommendation described at the beginning of this section is captured by $r^{\text{MB}*}$, defined as $r^{\text{MB}*}(s,b) = (p_S(s), p_B(b))$ for all $(s,b) \in \Theta_S \times \Theta_B$.

Let $V(q)$ denote the intermediary's ex ante expected payoff under pure mediation plan

[16] The buyer's manipulation could instead specify a response for each possible price. However, this generalization does not affect the results, as long as the analysis is restricted to acceptable CEs.

[17] Note that, in this game, every acceptable CE is a *sequential communication equilibrium* (SCE). Myerson (1986b) defines the notion of SCE and shows that a CE is an SCE if and only if it never recommends a *codominated* action to a player who has been truthful. Roughly speaking, an action is codominated if, *whenever* it is recommended with positive probability, at least one player could expect to gain by manipulation after being told to take that action. Given that $r^{\text{SO}*}$ recommends a maximum acceptable price equal to the reported type, the buyer clearly has no profitable deviation if he reports his type truthfully, implying that $r^{\text{SO}*}$ has no codominated action in its range. Hence, no $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$ recommends a codominated action to the buyer. Lemma 1 further implies that no acceptable CE recommends a codominated action to the seller either, thereby establishing the equivalence between CE and SCE (see footnote 19). A similar argument can be made for the mediated bargaining game.

[18] Hence, any price $p \in [p_S(s), p_B(b)]$ is mutually acceptable for a type-s seller and a type-b buyer. Since $h \le \bar{h}$ implies $0 < p_S(s_L) \le p_B(b_L) < p_S(s_H) \le p_B(b_H)$, a mutually acceptable price exists if and only if there is a gain from trade.

$(q, r^{\text{MB}*})$ when all players are honest and obedient:

$$V(q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)(\delta_S + \delta_B)q(s,b) \cdot \mathbf{1}_{\{p_S(s)\le q(s,b)\le p_B(b)\}}.$$

Likewise, the expected payoff for each trader $i \in \{S, B\}$ is given by

$$U_S^{\text{MB}}(q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[(1-\delta_S)q(s,b) - s] \cdot \mathbf{1}_{\{p_S(s)\le q(s,b)\le p_B(b)\}},$$
$$U_B^{\text{MB}}(q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[b - (1+\delta_B)q(s,b)] \cdot \mathbf{1}_{\{p_S(s)\le q(s,b)\le p_B(b)\}}.$$

The intermediary's manipulation is represented by $\sigma_I \colon \mathbb{R}_+ \to \mathbb{R}_+$. Let Σ_I denote the set of the intermediary's manipulations. If she manipulates $(q, r^{\text{MB}*})$ using $\sigma_I \in \Sigma_I$ while the traders are honest and obedient, her ex ante expected payoff is

$$V(\sigma_I \circ q) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)(\delta_S + \delta_B)\sigma_I(q(s,b)) \cdot \mathbf{1}_{\{p_S(s)\le \sigma_I(q(s,b))\le p_B(b)\}}.$$

The seller's manipulation is represented by $\sigma_S^{\text{MB}} = \left(\sigma_{S1}^{\text{MB}}, \sigma_{S2}^{\text{MB}}\right)$, where $\sigma_{S1}^{\text{MB}} \colon \Theta_S \to \Theta_S$ is a manipulation in report and $\sigma_{S2}^{\text{MB}} \colon \Theta_S \times \mathbb{R}_+ \to \mathbb{R}_+$ is a manipulation in *response*. Let Σ_S^{MB} denote the set of the seller's manipulations. If he manipulates $(q, r^{\text{MB}*})$ using $\sigma_S^{\text{MB}} \in \Sigma_S^{\text{MB}}$ while the other players are honest and obedient, his ex ante expected payoff is

$$U_S^{\text{MB}}\left(q \circ \sigma_S^{\text{MB}}\right) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)[(1-\delta_S)q(\tilde{s}(s), b) - s] \cdot \mathbf{1}_{\left\{\sigma_S^2(s, p_S(\tilde{s}(s)))\le q(\tilde{s}(s),b)\le p_B(b)\right\}},$$

where $\tilde{s}(s) \equiv \sigma_{S1}^{\text{MB}}(s)$. The buyer's manipulation is represented by $\sigma_B^{\text{MB}} = \left(\sigma_{B1}^{\text{MB}}, \sigma_{B2}^{\text{MB}}\right)$, where $\sigma_{B1}^{\text{MB}} \colon \Theta_B \to \Theta_B$ and $\sigma_{B2}^{\text{MB}} \colon \Theta_B \times \mathbb{R}_+ \to \mathbb{R}_+$. Let Σ_B^{MB} denote the set of the buyer's manipulations. If he manipulates $(q, r^{\text{MB}*})$ using $\sigma_B^{\text{MB}} \in \Sigma_B^{\text{MB}}$ while the other players are honest and obedient, his ex ante expected payoff is

$$U_B^{\text{MB}}\left(q \circ \sigma_B^{\text{MB}}\right) = \sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b)\left[b - (1+\delta_B)q\left(s, \tilde{b}(b)\right)\right] \cdot \mathbf{1}_{\left\{p_S(s)\le q\left(s,\tilde{b}(b)\right)\le \sigma_B^2\left(b, p_B\left(\tilde{b}(b)\right)\right)\right\}},$$

where $\tilde{b}(b) \equiv \sigma^{\text{MB}}_{B1}(b)$.

Definition 2. In the mediated bargaining game, a mediation plan $\mu^{\text{MB}} \in \Delta\big(Q \times \{r^{\text{MB}*}\}\big)$ is an *acceptable communication equilibrium* if the following two conditions hold:

1. For all $\sigma_I \in \Sigma_I$,
$$\sum_{q \in Q} \mu^{\text{MB}}(q) V(q) \geq \sum_{q \in Q} \mu(q) V(\sigma_I \circ q); \tag{2.2}$$

2. For all $i \in \{S, B\}$ and all $\sigma_i^{\text{MB}} \in \Sigma_i^{\text{MB}}$,
$$\sum_{q \in Q} \mu^{\text{MB}}(q) U_i^{\text{MB}}(q) \geq \sum_{q \in Q} \mu(q) U_i^{\text{MB}}\big(q \circ \sigma_i^{\text{MB}}\big). \tag{2.3}$$

Remark 1. Strictly speaking, the communication revelation principle is established only for finite games. Moreover, its literal application requires modifying the timing so that a communication device recommends actions *sequentially*—that is, recommending a price at Time 4 and a response after Time 5. As long as the analysis is restricted to acceptable CEs, however, I can obtain qualitatively the same result by applying the communication revelation principle to a modified game in which the set of possible prices is discretized and recommendations are given sequentially. For clarity and brevity, I adopt the current formulation throughout the paper.

Remark 2. In both games, the players' ex ante incentive compatibility (IC) constraint—namely, (2.1) in the seller-offer bargaining game and (2.2) and (2.3) in the mediated bargaining game—are satisfied if and only if the corresponding interim IC constraints are satisfied; that is, no player has an incentive to deviate after learning their type or receiving a recommendation. I thus focus on the interim IC constraints hereafter, as they are more tractable.

Intuitively, if a player has a profitable manipulation in some interim scenario, then I can construct an ex ante manipulation that prescribes deviation only in that scenario and follows the recommendation otherwise. Such a manipulation is clearly profitable ex ante. Conversely, if a player has a profitable ex ante manipulation, then there must exist at least one interim scenario in which the player gains from deviating. This argument establishes the equivalence between ex ante and interim incentive compatibility.

3 Preliminary Results

In both games, the player who proposes a price has a payoff that increases with the trading price. Given the structure of the response recommendations under consideration—either $r^{\mathrm{SO}*}$ or $r^{\mathrm{MB}*}$—this provides an incentive to offer the maximum acceptable price to the buyer whenever there is a gain from trade. Since the buyer has only two possible types, any acceptable CE involves at most two different prices for type profiles with gains from trade, which substantially simplifies the players' IC constraints. This section formalizes this observation and derives the players' IC constraints in both games.

3.1 Seller-offer bargaining game

As discussed above, any acceptable CE recommends that the seller offer either b_L or b_H. In particular, no acceptable CE recommends the price b_L to the high-type seller, since it is not acceptable for him. This leads to the following necessary condition for acceptable CE.[19]

Lemma 1. *A mediation plan $\mu^{\mathrm{SO}} \in \Delta\big(Q \times \{r^{\mathrm{SO}*}\}\big)$ is an acceptable CE only if it always recommends the price b_H when the reported types are (s_H, b_H) and recommends either b_L or b_H when the seller reports s_L. That is, for all $q \in \operatorname{supp}\big(\mu^{\mathrm{SO}}\big)$,*

$$\begin{aligned} &q(s_H, b_H) = b_H, \\ &q(s_L, b) \in \{b_L, b_H\} \textit{ for all } b \in \Theta_B. \end{aligned} \tag{3.1}$$

Proof. I prove the contraposition. Consider a mediation plan $\mu^{\mathrm{SO}} \in \Delta\big(Q \times \{r^{\mathrm{SO}*}\}\big)$ that violates (3.1) for some $q \in \operatorname{supp}\big(\mu^{\mathrm{SO}}\big)$. I show that the seller has a profitable manipulation. If type s seller reports $\tilde{s}$ and receives a recommendation p, his posterior belief that the buyer is high type is

$$\nu^{\mathrm{SO}}(b_H \mid \tilde{s}, p) = \frac{\pi_B \sum_{q:\ q(\tilde{s}, b_H) = p} \mu^{\mathrm{SO}}(q)}{\pi_B \sum_{q:\ q(\tilde{s}, b_H) = p} \mu^{\mathrm{SO}}(q) + (1 - \pi_B) \sum_{q:\ q(\tilde{s}, b_L) = p} \mu^{\mathrm{SO}}(q)}.$$

[19] As can be inferred from the proof of the lemma, when the seller is recommended b_L, he cannot gain by deviating if he believes that the buyer is sufficiently likely to be low type. Similarly, when recommended b_H, he cannot gain by deviating either if he believes that the buyer is sufficiently likely to be high type. This implies that the actions b_L and b_H are not codominated. Therefore, no acceptable CE recommends a codominated action to the seller.

Assuming the buyer is honest and obedient, the seller's expected payoff from offering $\tilde{p}$ is

$$\begin{cases} \tilde{p} - s & \text{if } \tilde{p} \in [0, b_L]; \\ \nu^{\text{SO}}(b_H \mid \tilde{s}, p)(\tilde{p} - s) & \text{if } \tilde{p} \in (b_L, b_H]; \\ 0 & \text{if } \tilde{p} \in (b_H, +\infty). \end{cases}$$

First, suppose $q(s_H, b_H) = p \neq b_H$. If the high-type seller reports his type truthfully and receives a recommendation p, his posterior is $\nu^{\text{SO}}(b_H \mid s_H, p) > 0$. He would thus prefer to offer b_H. Next, suppose $q(s_L, b) = p \notin \{b_L, b_H\}$ for some $b \in \Theta_B$. If the low-type seller reports his type truthfully and receives a recommendation p, his posterior is $\nu^{\text{SO}}(b_H \mid s_L, p) \geq 0$. If $p \in (b_L, b_H)$ and $\nu^{\text{SO}}(b_H \mid s_L, p) > 0$, he would prefer to offer b_H. Otherwise, he would prefer to offer b_L. □

To further narrow down the candidates for acceptable CEs, consider the recommendation when the reported types are (s_H, b_L). Given any mediation plan $\mu^{\text{SO}} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$, construct a modified mediation plan $\mu^{\text{SO}*}$ as follows. For each $q \in \text{supp}\big(\mu^{\text{SO}}\big)$, define $q^{\text{SO}*} \in Q$ by

$$q^{\text{SO}*}(s, b) = \begin{cases} q(s, b) & \text{if } (s, b) \neq (s_H, b_L); \\ b_H & \text{if } (s, b) = (s_H, b_L). \end{cases}$$

Define $\mu^{\text{SO}*} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$ by setting $\mu^{\text{SO}*}\big(q^{\text{SO}*}\big) = \mu^{\text{SO}}(q)$. That is, $\mu^{\text{SO}*}$ is obtained by modifying each q in the support of μ^{SO} so that the recommendation for (s_H, b_L) is replaced with b_H, while leaving all other entries unchanged.

Two properties of $\mu^{\text{SO}*}$ make it without loss of generality to focus on the set of such mediation plans. First, each $q^{\text{SO}*}$ eliminates profitable *double deviations*—misreporting followed by disobedience—by the low-type seller and high-type buyer, which may exist under q. Thus, the players' IC constraints under μ^{SO} are weakly more stringent than those under $\mu^{\text{SO}*}$. Second, μ^{SO} and $\mu^{\text{SO}*}$ yield the same ex ante expected social surplus when all players are honest and obedient.[20] This leads to the following lemma.

[20]Since the trader pair (s_H, b_L) never trades in equilibrium, modifying the recommendation for this pair does not affect the expected social surplus.

Lemma 2. *For any mediation plan* $\mu^{\text{SO}} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$*, if* μ^{SO} *is an acceptable CE, then so is* $\mu^{\text{SO}*}$*. Moreover,* μ^{SO} *and* $\mu^{\text{SO}*}$ *yield the same ex ante expected social surplus when all players are honest and obedient.*

Proof. Given Lemma 1, consider an arbitrary mediation plan $\mu^{\text{SO}} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$ that satisfies (3.1) for all $q \in \text{supp}(\mu^{\text{SO}})$. By construction, both traders obtain the same expected payoff under μ^{SO} and $\mu^{\text{SO}*}$ if they are honest and obedient. Hence, it suffices to show that μ^{SO} allows more room for profitable manipulation by both players than $\mu^{\text{SO}*}$.

Let $x_{\text{HL}}^{\text{SO}}$ and $x_{\text{LL}}^{\text{SO}}$ denote the total probabilities under μ^{SO} that the price b_H is recommended for the report (s_H, b_L) and (s_L, b_L), respectively:

$$x_{\text{HL}}^{\text{SO}} = \sum_{q:\ q(s_H,b_L)=b_H} \mu^{\text{SO}}(q),$$
$$x_{\text{LL}}^{\text{SO}} = \sum_{q:\ q(s_L,b_L)=b_H} \mu^{\text{SO}}(q).$$

Note that the latter probability coincides under μ^{SO} and $\mu^{\text{SO}*}$ because the recommendation for (s_L, b_L) remains unchanged.

As the difference between μ^{SO} and $\mu^{\text{SO}*}$ does not affect the IC constraints of the high-type seller and the low-type buyer, it remains to show that those of the low-type seller and the high-type buyer are (weakly) more stringent under μ^{SO} than under $\mu^{\text{SO}*}$.

Low-type seller. Under μ^{SO}, if the low-type seller misreports his type and receives a recommendation $p \neq b_H$, he learns that the buyer is low type. Hence, his expected payoff is maximized by either (i) offering b_H when the recommended price is b_H, and offering b_L otherwise; or (ii) offering b_L regardless of the recommendation. His expected payoff from misreporting is thus at most

$$\max\big\{\pi_B(b_H - s_L) + (1-\pi_B)(1-x_{\text{HL}}^{\text{SO}})(b_L - s_L), b_L - s_L\big\}.$$

By contrast, under $\mu^{\text{SO}*}$, his expected payoff from misreporting is at most

$$\max\{\pi_B(b_H - s_L), b_L - s_L\} \tag{3.2}$$

because he always receives the recommendation b_H, and thus his expected payoff is maximized by offering either b_H or b_L. Therefore, his IC constraint is (weakly) more stringent under μ^{SO} than under $\mu^{\mathrm{SO}*}$.

High-type buyer. Under μ^{SO}, if the high-type buyer misreports his type, he can obtain a payoff of $b_H - b_L$ when the seller is low type and offers the price b_L, which occurs with probability $(1-\pi_S)(1-x^{\mathrm{SO}}_{\mathrm{LL}})$. Moreover, if the seller is high-type, he may offer a price $p \in [s_H, b_H)$, which the buyer can accept to obtain a strictly positive payoff. By contrast, under $\mu^{\mathrm{SO}*}$, such prices are never offered by the high-type seller if he is obedient. Formally, the high-type buyer's expected payoff from misreporting under μ^{SO} is at most

$$(1-\pi_S)(1-x^{\mathrm{SO}}_{\mathrm{LL}})(b_H - b_L) + \pi_S \sum_{p\in[s_H,b_H)} \sum_{q:\ q(s_H,b_L)=p} \mu^{\mathrm{SO}}(q)(b_H - p),$$

while under $\mu^{\mathrm{SO}*}$, it is

$$(1-\pi_S)(1-x^{\mathrm{SO}}_{\mathrm{LL}})(b_H - b_L). \tag{3.3}$$

Therefore, his IC constraint is also (weakly) more stringent under μ^{SO} than under $\mu^{\mathrm{SO}*}$.

Finally, consider the ex ante expected social surplus. Since the trading price cancels out, the realized social surplus from a trade between types (s, b) is simply $b - s$. Hence, the ex ante expected social surplus under μ^{SO} is given by

$$\begin{aligned}
&\sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b) \sum_{q\in Q} \mu^{\mathrm{SO}}(q)(b-s)\cdot \mathbf{1}_{\{q(s,b)\le b\}} \\
&= \pi_S\pi_B(b_H - s_H) + (1-\pi_S)\pi_B(b_H - s_L) + (1-\pi_S)(1-\pi_B)(1-x^{\mathrm{SO}}_{\mathrm{LL}})(b_L - s_L).
\end{aligned}$$

This expression does not depend on $x^{\mathrm{SO}}_{\mathrm{HL}}$, the only difference between μ^{SO} and $\mu^{\mathrm{SO}*}$. Therefore, the expected social surplus coincides under the two mediation plans. □

3.1.1 IC constraints

By Lemma 2, it suffices to focus on $\mu^{\mathrm{SO}*}$, which is constructed from some $\mu^{\mathrm{SO}} \in \Delta\big(Q \times \{r^{\mathrm{SO}*}\}\big)$ that satisfies (3.1) for all $q \in \mathrm{supp}\big(\mu^{\mathrm{SO}}\big)$. As in the proof of Lemma 2, define $x^{\mathrm{SO}}_{\mathrm{LH}}$ as the total probability under μ^{SO} that the price b_H is recommended for the report (s_L, b_H):

$$x^{\text{SO}}_{\text{LH}} = \sum_{q:\, q(s_L, b_H) = b_H} \mu^{\text{SO}}(q).$$

As in the case of $x^{\text{SO}}_{\text{LL}}$, this probability coincides under μ^{SO} and $\mu^{\text{SO}*}$ because the recommendation for (s_L, b_H) remains unchanged. Then, the players' IC constraints under $\mu^{\text{SO}*}$ can be expressed as linear inequalities in $x^{\text{SO}}_{\text{LH}}$ and $x^{\text{SO}}_{\text{LL}}$, as shown below.

High-type seller. If the high-type seller is honest and obedient, he obtains a positive payoff only when the buyer is also high type. In that case, he offers the price b_H and obtains $b_H - s_H$, yielding an expected payoff of $\pi_B(b_H - s_H)$. He cannot do better if he reports his type truthfully. If he misreports his type, offering b_H remains his best option—yielding the same expected payoff $\pi_B(b_H - s_H)$—because trading with the low-type buyer never yields a positive payoff. Thus, his IC constraint is trivially satisfied.

Low-type seller. If the low-type seller is honest and obedient, he obtains $b_H - s_L$ when the buyer is high type and the recommended price is b_H, and obtains $b_L - s_L$ when the recommended price is b_L, regardless of the buyer's type. Hence, his expected payoff is

$$\pi_B\left[x^{\text{SO}}_{\text{LH}}(b_H - s_L) + (1 - x^{\text{SO}}_{\text{LH}})(b_L - s_L)\right] + (1 - \pi_B)(1 - x^{\text{SO}}_{\text{LL}})(b_L - s_L).$$

Combined with (3.2), he has no incentive to misreport his type if

$$\begin{aligned} &\pi_B\left[x^{\text{SO}}_{\text{LH}}(b_H - s_L) + (1 - x^{\text{SO}}_{\text{LH}})(b_L - s_L)\right] + (1 - \pi_B)(1 - x^{\text{SO}}_{\text{LL}})(b_L - s_L) \\ &\quad \geq \max\{\pi_B(b_H - s_L), b_L - s_L\}. \end{aligned} \tag{SO-IC$_{s_L}$}$$

Now, suppose that he reports his type truthfully and receives the recommendation b_H. His posterior belief that the buyer is high type is $\frac{\pi_B x^{\text{SO}}_{\text{LH}}}{\pi_B x^{\text{SO}}_{\text{LH}}+(1-\pi_B)x^{\text{SO}}_{\text{LL}}}$. Hence, he follows the recommendation if [21]

$$\frac{\pi_B x^{\text{SO}}_{\text{LH}}}{\pi_B x^{\text{SO}}_{\text{LH}} + (1-\pi_B)x^{\text{SO}}_{\text{LL}}}(b_H - s_L) \geq b_L - s_L$$
$$\iff \pi_B x^{\text{SO}}_{\text{LH}}(b_H - b_L) \geq (1-\pi_B)x^{\text{SO}}_{\text{LL}}(b_L - s_L). \qquad (\text{SO-IC}_{s_L}\text{-1})$$

Similarly, if he receives the recommendation b_L, his posterior is $\frac{\pi_B(1-x^{\text{SO}}_{\text{LH}})}{\pi_B(1-x^{\text{SO}}_{\text{LH}})+(1-\pi_B)(1-x^{\text{SO}}_{\text{LL}})}$. Hence, he follows the recommendation if

$$b_L - s_L \geq \frac{\pi_B(1-x^{\text{SO}}_{\text{LH}})}{\pi_B(1-x^{\text{SO}}_{\text{LH}}) + (1-\pi_B)(1-x^{\text{SO}}_{\text{LL}})}(b_H - s_L)$$
$$\iff (1-\pi_B)(1-x^{\text{SO}}_{\text{LL}})(b_L - s_L) \geq \pi_B(1-x^{\text{SO}}_{\text{LH}})(b_H - b_L). \qquad (\text{SO-IC}_{s_L}\text{-2})$$

Note that (SO-IC_{s_L}) is equivalent to (SO-IC_{s_L}-1) if $\pi_B(b_H - s_L) < b_L - s_L$, and to ($\text{SO-IC}_{s_L}$-2) otherwise.

High-type buyer. If the high-type buyer is honest and obedient, he obtains a positive payoff of $b_H - b_L$ only when the seller is low type and offers b_L, yielding an expected payoff of $(1-\pi_S)(1-x^{\text{SO}}_{\text{LH}})(b_H - b_L)$. Combined with (3.3), his IC constraint is

$$(1-\pi_S)(1-x^{\text{SO}}_{\text{LH}})(b_H - b_L) \geq (1-\pi_S)(1-x^{\text{SO}}_{\text{LL}})(b_H - b_L) \iff x^{\text{SO}}_{\text{LL}} \geq x^{\text{SO}}_{\text{LH}}. \qquad (\text{SO-IC}_{b_H})$$

Low-type buyer. The low-type buyer obtains zero expected payoff if he is honest and obedient. As the seller only offers b_L or b_H, no manipulation yields him a positive expected payoff. Thus, his IC constraint is trivially satisfied.

Therefore, any $\mu^{\text{SO}*}$ satisfying (SO-IC_{s_L}) and (SO-IC_{b_H}) is an acceptable CE.

[21] As discussed in the proof of Lemma 2, it suffices to show that he prefers offering b_H over b_L. Similar argument applies to the recommendation b_L.

3.2 Mediated bargaining game

For the mediated bargaining game, analogs of Lemmas 1 and 2 can be established. Their proofs closely follow the original ones and are omitted in the main text.

First, any acceptable CE recommends that the intermediary offer either $p_B(b_L)$ or $p_B(b_H)$ whenever there is a gain from trade.[22] The key difference from the seller-offer bargaining game is that the recommendation for the report (s_H, b_H) is not necessarily the high price $p_B(b_H)$, since the intermediary does not know the seller's type and may have an incentive to offer $p_B(b_L)$.

Lemma 3. *A mediation plan $\mu^{\mathrm{MB}} \in \Delta\big(Q \times \{r^{\mathrm{MB}*}\}\big)$ is an acceptable CE only if it always recommends either $p_B(b_L)$ or $p_B(b_H)$ whenever there is a gain from trade. That is, for all $q \in \mathrm{supp}(\mu^{\mathrm{MB}})$,*

$$q(s,b) \in \{p_B(b_L), p_B(b_H)\} \quad \textit{for all } (s,b) \in \Theta_S \times \Theta_B \setminus \{(s_H, b_L)\}. \tag{3.4}$$

Proof. See Appendix A. □

Intuitively, when the traders follow $r^{\mathrm{MB}*}$, trade occurs if and only if the intermediary offers a mutually acceptable price. Thus, she clearly has an incentive to offer the maximum acceptable price to the buyer—either $p_B(b_L)$ or $p_B(b_H)$. Therefore, if a mediation plan were to recommend a price other than these two, the intermediary would have an incentive to deviate from the recommendation.

As in the seller-offer bargaining game, it is without loss of generality to focus on the recommendations to (s_H, b_L) that eliminate the possibility of profitable double deviations. Specifically, fix some $p_{\mathrm{HL}} \in (p_B(b_H), +\infty)$, and for any mediation plan $\mu^{\mathrm{MB}} \in \Delta\big(Q \times \{r^{\mathrm{MB}*}\}\big)$, construct a modified mediation plan $\mu^{\mathrm{MB}*}$ as follows. For each $q \in \mathrm{supp}(\mu^{\mathrm{MB}})$, define $q^{\mathrm{MB}*} \in Q$ by

$$q^{\mathrm{MB}*}(s,b) = \begin{cases} q(s,b) & \text{if } (s,b) \neq (s_H, b_L); \\ p_{\mathrm{HL}} & \text{if } (s,b) = (s_H, b_L). \end{cases}$$

Define $\mu^{\mathrm{MB}*}$ by setting $\mu^{\mathrm{MB}*}(q^{\mathrm{MB}*}) = \mu^{\mathrm{MB}}(q)$. That is, $\mu^{\mathrm{MB}*}$ is obtained by a similar modification as in the case of $\mu^{\mathrm{SO}*}$, except that the recommendation for (s_H, b_L) is now replaced

[22] Recall that $p_B(b)$ is the price at which a type-b buyer breaks even.

with p_{HL}, which is not mutually acceptable for any trader pair. Compared to the original mediation plan, the modified plan $\mu^{\mathrm{MB}*}$ entails weakly less stringent IC constraints and yields the same ex ante expected social surplus when the players are honest and obedient. This leads to the following lemma.

Lemma 4. *For any mediation plan $\mu^{\mathrm{MB}} \in \Delta\left(Q \times \left\{r^{\mathrm{MB}*}\right\}\right)$, if μ^{MB} is an acceptable CE, then so is $\mu^{\mathrm{MB}*}$. Moreover, μ^{MB} and $\mu^{\mathrm{MB}*}$ yield the same ex ante expected social surplus when all players are honest and obedient.*

Proof. See Appendix B. □

3.2.1 IC constraints

By Lemma 4, it suffices to focus on $\mu^{\mathrm{MB}*}$, which is constructed from some $\mu^{\mathrm{MB}} \in \Delta\left(Q \times \left\{r^{\mathrm{MB}*}\right\}\right)$ that satisfies (3.4) for all $q \in \mathrm{supp}\left(\mu^{\mathrm{MB}}\right)$. As in the seller offer bargaining game, let $x_{\mathrm{HH}}^{\mathrm{MB}}$, $x_{\mathrm{LH}}^{\mathrm{MB}}$, and $x_{\mathrm{LL}}^{\mathrm{MB}}$ denote the total probability under μ^{MB} that the price $p_B(b_H)$ is recommended for the report (s_H, b_H), (s_L, b_H), and (s_L, b_L), respectively:

$$
\begin{aligned}
x_{\mathrm{HH}}^{\mathrm{MB}} &= \sum_{q:\, q(s_H, b_H) = p_B(b_H)} \mu^{\mathrm{MB}}(q), \\
x_{\mathrm{LH}}^{\mathrm{MB}} &= \sum_{q:\, q(s_L, b_H) = p_B(b_H)} \mu^{\mathrm{MB}}(q), \\
x_{\mathrm{LL}}^{\mathrm{MB}} &= \sum_{q:\, q(s_L, b_L) = p_B(b_H)} \mu^{\mathrm{MB}}(q).
\end{aligned}
$$

These probabilities coincide under μ^{MB} and $\mu^{\mathrm{MB}*}$ because the relevant recommendations remain unchanged. Then, the players' IC constraints under $\mu^{\mathrm{MB}*}$ can be expressed as linear inequalities in $x_{\mathrm{HH}}^{\mathrm{MB}}$, $x_{\mathrm{LH}}^{\mathrm{MB}}$, and $x_{\mathrm{LL}}^{\mathrm{MB}}$, as shown below.

High-type seller. If the high-type seller is honest and obedient, he obtains a payoff of $(1-\delta_S)p_B(b_H) - s_H = \frac{b_H}{h} - s_H$ when the buyer is also high type and the price $p_B(b_H)$ is offered. In all other cases, he cannot obtain a positive payoff. If he misreports his type, he can obtain the same payoff only in the same scenario. Thus, he has no incentive to misreport if [23]

$$\pi_B x_{\text{HH}}^{\text{MB}}\left(\frac{b_H}{h} - s_H\right) \geq \pi_B x_{\text{LH}}^{\text{MB}}\left(\frac{b_H}{h} - s_H\right) \iff x_{\text{HH}}^{\text{MB}} \geq x_{\text{LH}}^{\text{MB}}. \qquad (\text{MB-IC}_{s_H})$$

Low-type seller. If the low-type seller is honest and obedient, he obtains a payoff of $(1 - \delta_S)p_B(b_H) - s_L = \frac{b_H}{h} - s_L$ when the buyer is high type and the price $p_B(b_H)$ is offered, and $(1 - \delta_S)p_B(b_L) - s_L = \frac{b_L}{h} - s_L$ when the price $p_B(b_L)$ is offered, regardless of the buyer's type. If he misreports his type, he can still obtain the same payoffs when these prices are offered, but they may be offered only when the buyer is high type. Hence, his expected payoff from misreporting under $\mu^{\text{MB}*}$ is at most

$$\pi_B\left[x_{\text{HH}}^{\text{MB}}\frac{b_H}{h} + \left(1 - x_{\text{HH}}^{\text{MB}}\right)\frac{b_L}{h} - s_L\right].$$

Thus, he has no incentive to misreport if

$$\begin{aligned}
&\pi_B\left[x_{\text{LH}}^{\text{MB}}\frac{b_H}{h} + \left(1 - x_{\text{LH}}^{\text{MB}}\right)\frac{b_L}{h} - s_L\right] + (1 - \pi_B)(1 - x_{\text{LL}}^{\text{MB}})\left(\frac{b_L}{h} - s_L\right) \\
&\geq \pi_B\left[x_{\text{HH}}^{\text{MB}}\frac{b_H}{h} + \left(1 - x_{\text{HH}}^{\text{MB}}\right)\frac{b_L}{h} - s_L\right] \\
\iff &(1 - \pi_B)(1 - x_{\text{LL}}^{\text{MB}})(b_L - hs_L) \geq \pi_B(x_{\text{HH}}^{\text{MB}} - x_{\text{LH}}^{\text{MB}})(b_H - b_L). \qquad (\text{MB-IC}_{s_L})
\end{aligned}$$

High-type buyer. If the high-type buyer is honest and obedient, he obtains a payoff of $b_H - (1 + \delta_B)p_B(b_H) = b_H - b_L$ when the seller is low type and price $p_B(b_L)$ is offered. In all other cases, he cannot obtain a positive payoff. If he misreports his type, he can obtain the same payoff only in the same scenario. Hence, his expected payoff from misreporting under $\mu^{\text{MB}*}$ is at most

$$(1 - \pi_S)(1 - x_{\text{LL}}^{\text{MB}})(b_H - b_L).$$

Thus, he has no incentive to misreport if

$$(1 - \pi_S)(1 - x_{\text{LH}}^{\text{MB}})(b_H - b_L) \geq (1 - \pi_S)(1 - x_{\text{LL}}^{\text{MB}})(b_H - b_L) \iff x_{\text{LL}}^{\text{MB}} \geq x_{\text{LH}}^{\text{MB}}. \quad (\text{MB-IC}_{b_H})$$

[23] Recall that if the traders report their type truthfully, they cannot do better than following the recommendation $r^{\text{MB}*}$. Hence, it suffices to show that they have no incentive to misreport their type.

Low-type buyer. The low-type buyer obtains zero expected payoff if he is honest and obedient. As the intermediary only offers $p_B(b_L)$ or $p_B(b_H)$, no manipulation yields him a positive expected payoff. Thus, his IC constraint is trivially satisfied.

Intermediary. If the intermediary receives the recommendation $p_B(b_H)$, her posterior beliefs about the traders' types are as follows:

$$\begin{aligned}
\nu^{\text{MB}}(s_H, b_H \mid p_B(b_H)) &= \frac{\pi_S \pi_B x_{\text{HH}}^{\text{MB}}}{D_{b_H}}, \\
\nu^{\text{MB}}(s_L, b_H \mid p_B(b_H)) &= \frac{(1-\pi_S)\pi_B x_{\text{LH}}^{\text{MB}}}{D_{b_H}}, \\
\nu^{\text{MB}}(s_L, b_L \mid p_B(b_H)) &= \frac{(1-\pi_S)(1-\pi_B) x_{\text{LL}}^{\text{MB}}}{D_{b_H}}, \\
\nu^{\text{MB}}(s_H, b_L \mid p_B(b_H)) &= 0,
\end{aligned}$$

where $D_{b_H} = \pi_S \pi_B x_{\text{HH}}^{\text{MB}} + (1-\pi_S)\pi_B x_{\text{LH}}^{\text{MB}} + (1-\pi_S)(1-\pi_B) x_{\text{LL}}^{\text{MB}}$ is the total probability that the price $p_B(b_H)$ is recommended. She follows the recommendation if she prefers offering $p_B(b_H)$ over $p_B(b_L)$:

$$\begin{aligned}
&\left[\nu^{\text{MB}}(s_H, b_H \mid p_B(b_H)) + \nu^{\text{MB}}(s_L, b_H \mid p_B(b_H))\right](\delta_S + \delta_B) p_B(b_H) \\
&\quad \geq \left[\nu^{\text{MB}}(s_L, b_H \mid p_B(b_H)) + \nu^{\text{MB}}(s_L, b_L \mid p_B(b_H))\right](\delta_S + \delta_B) p_B(b_L) \\
\iff &\pi_S \pi_B x_{\text{HH}}^{\text{MB}} b_H + (1-\pi_S)\pi_B x_{\text{LH}}^{\text{MB}} (b_H - b_L) \geq (1-\pi_S)(1-\pi_B) x_{\text{LL}}^{\text{MB}} b_L. \qquad (\text{IC}_{\text{Int}}\text{-1})
\end{aligned}$$

Similarly, if she receives the recommendation $p_B(b_L)$, her posterior beliefs are:

$$\begin{aligned}
\nu^{\text{MB}}(s_H, b_H \mid p_B(b_L)) &= \frac{\pi_S \pi_B (1 - x_{\text{HH}}^{\text{MB}})}{D_{b_L}}, \\
\nu^{\text{MB}}(s_L, b_H \mid p_B(b_L)) &= \frac{(1-\pi_S)\pi_B (1 - x_{\text{LH}}^{\text{MB}})}{D_{b_L}}, \\
\nu^{\text{MB}}(s_L, b_L \mid p_B(b_L)) &= \frac{(1-\pi_S)(1-\pi_B)(1 - x_{\text{LL}}^{\text{MB}})}{D_{b_L}}, \\
\nu^{\text{MB}}(s_H, b_L \mid p_B(b_L)) &= 0.
\end{aligned}$$

where $D_{b_L} = \pi_S \pi_B (1 - x_{\text{HH}}^{\text{MB}}) + (1-\pi_S)\pi_B (1 - x_{\text{LH}}^{\text{MB}}) + (1-\pi_S)(1-\pi_B)(1 - x_{\text{LL}}^{\text{MB}})$ is the total probability that the price $p_B(b_L)$ is recommended. She follows the recommendation if she prefers offering $p_B(b_L)$ over $p_B(b_H)$:

$$\begin{aligned}
&\left[v^{\mathrm{MB}}(s_L, b_H \mid p_B(b_L)) + v^{\mathrm{MB}}(s_L, b_L \mid p_B(b_L))\right](\delta_S + \delta_B)p_B(b_L) \\
&\quad \geq \left[v^{\mathrm{MB}}(s_H, b_H \mid p_B(b_L)) + v^{\mathrm{MB}}(s_L, b_H \mid p_B(b_L))\right](\delta_S + \delta_B)p_B(b_H) \\
\iff &\pi_S\pi_B x^{\mathrm{MB}}_{\mathrm{HH}} b_H + (1-\pi_S)\pi_B x^{\mathrm{MB}}_{\mathrm{LH}}(b_H - b_L) \geq (1-\pi_S)(1-\pi_B)x^{\mathrm{MB}}_{\mathrm{LL}} b_L + J, \qquad (\mathrm{IC}_{\mathrm{Int}}\text{-}2)
\end{aligned}$$

where J is defined as

$$J = \pi_S\pi_B b_H + (1-\pi_S)\pi_B(b_H - b_L) - (1-\pi_S)(1-\pi_B)b_L.$$

Thus, if $J \geq 0$, then ($\mathrm{IC}_{\mathrm{Int}}$-2) implies ($\mathrm{IC}_{\mathrm{Int}}$-1), and vice versa. Note also that $J \geq 0$ is equivalent to ($\mathrm{IC}_{\mathrm{Int}}$-1) under $x^{\mathrm{MB}}_{\mathrm{HH}} = x^{\mathrm{MB}}_{\mathrm{LH}} = x^{\mathrm{MB}}_{\mathrm{LL}} = 1$. That is, $J \geq 0$ implies that the intermediary follows the recommendation $p_B(b_H)$ when her belief stays at the prior.

Finally, if she receives the recommendation p_{HL}, she learns that the traders' types are (s_H, b_L). Since there is no mutually acceptable price in this case, she has no incentive to deviate from the recommendation.

Therefore, any $\mu^{\mathrm{MB}*}$ satisfying (MB-IC$_{s_H}$) (MB-IC$_{s_L}$), (MB-IC$_{b_H}$), ($\mathrm{IC}_{\mathrm{Int}}$-1), and ($\mathrm{IC}_{\mathrm{Int}}$-2) is an acceptable CE.

4 Bound on the Ex Ante Expected Social Surplus

As the IC constraints derived in Sections 3.1.1 and 3.2.1, as well as the ex ante expected social surplus, are linear in $x^{\mathrm{SO}} = (x^{\mathrm{SO}}_{\mathrm{LH}}, x^{\mathrm{SO}}_{\mathrm{LL}})$ or $x^{\mathrm{MB}} = (x^{\mathrm{MB}}_{\mathrm{HH}}, x^{\mathrm{MB}}_{\mathrm{LH}}, x^{\mathrm{MB}}_{\mathrm{LL}})$, the upper bound on the expected social surplus achievable in acceptable CEs can be computed by solving the corresponding linear program.

In this section, I focus directly on the second-best (SB) case, where *ex post efficiency* cannot be achieved in any acceptable CE, and omit a detailed analysis of the conditions under which ex post efficiency can be achieved, as this is tangential to the main message of the paper. Instead, I briefly outline the intuition here. First, the definition of *ex post efficiency*—that is, trade occurs if and only if the buyer has a higher valuation—immediately determines the price recommendation for the report (s_L, b_L) (and also for (s_H, b_H) in the mediated bargaining game), since only one of the two possible on-path prices—b_L and b_H in the seller-offer bargaining game, and $p_B(b_L)$ and $p_B(b_H)$ in the mediated bargaining game—is mutually acceptable. The low price (b_L or $p_B(b_L)$) must be recommended for (s_L, b_L), and the high price $p_B(b_H)$ for (s_H, b_H). It then follows that the recommendation for (s_L, b_H) must always be the low price; otherwise, the high-type buyer would have an incentive to misreport his type to obtain it.[24]

Having identified the candidate mediation plans, it remains to find the conditions under which they constitute an acceptable CE. In the seller-offer bargaining game, it is straightforward to see that the low-type seller must believe that the buyer is sufficiently likely to be low type, who would reject the high price b_H. Otherwise, he would deviate and offer b_H even when recommended the low price b_L. In the mediated bargaining game, the only nontrivial IC constraints are those of the low-type seller and the intermediary. The low-type seller can obtain the low price $p_B(b_L)$ by being honest and obedient. However, if he misreports his type, he obtains the high price $p_B(b_H)$ when the buyer is high type. Similarly, when the intermediary is recommended the low price, trade occurs with certainty if she follows the recommendation. However, if she deviates and offers the high price instead, she obtains a higher commission when the buyer is high type. To deter such misreporting and deviation, both the seller and the intermediary must believe that the buyer is sufficiently likely to be low type. This requirement imposes an upper bound on the prior probability π_B that the buyer is high type. In the mediated bargaining game, this upper bound naturally depends on the ratio $h = \frac{1+\delta_B}{1-\delta_S}$. A formal analysis (omitted for brevity) leads to the following proposition.[25]

Proposition 1. *There exists an acceptable CE that achieves ex post efficiency under the following*

[24] That is, the candidate mediation plans are $x^{\mathrm{SO}} = (0, 0)$ in the seller-offer bargaining game and $x^{\mathrm{MB}} = (1, 0, 0)$ in the mediated bargaining game. Substituting these values into the respective IC constraints confirms the argument in the next paragraph and establishes the proposition that follows.

[25] I can also consider a purely benevolent intermediary who takes no commissions and aims to maximize, say, the expected trade surplus. In this case, ex post efficiency can be achieved under a weaker condition, suggesting that even slight commissions may undermine the possibility of achieving ex post efficiency. This condition coincides with that identified by Matsuo (1989) for the existence of a Bayesian incentive-compatible, individually rational, and ex post efficient trading mechanism in a binary-valuation version of the Myerson and Satterthwaite (1983) setting. It thus follows that if the intermediary is unbiased, neither commitment nor enforcement power is necessary to achieve ex post efficiency. This is because, regarding commitment, a benevolent intermediary can be incentivized to offer any mutually acceptable price, thereby effectively acquiring commitment. As for enforcement, ex post IR constraint is satisfied even in Matsuo (1989), so no enforcement power is required.

conditions:

1. *in the seller-offer bargaining game if and only if* $\pi_B \leq \frac{b_L - s_L}{b_H - s_L}$*; and*

2. *in the mediated bargaining game if and only if* $\pi_B \leq \frac{b_L - hs_L}{b_H - hs_L}$.

Note that the condition $\pi_B \leq \frac{b_L - hs_L}{b_H - hs_L}$ can also be interpreted as imposing an upper bound on the ratio h, denoted by $\bar{h}^{\text{Eff}}$:[26]

$$\pi_B \leq \frac{b_L - hs_L}{b_H - hs_L} \iff h \leq \frac{b_L}{s_L} - \frac{\pi_B(b_H - b_L)}{(1 - \pi_B)s_L} \equiv \bar{h}^{\text{Eff}}.$$

Intuitively, for a fixed π_B, an increase in h reduces the seller's payoff in all circumstances through a decrease in the equilibrium prices $p_B(b_L)$ and $p_B(b_H)$. However, if π_B is sufficiently small (specifically, if $\pi_B \leq \frac{b_L}{b_H}$), his equilibrium payoff decreases faster than his maximum payoff from misreporting, making deviation more attractive. Hence, the ratio h must be small enough to deter the low-type seller from misreporting. It is straightforward to verify that $\bar{h}^{\text{Eff}} \leq 1 \Leftrightarrow \pi_B \geq \frac{b_L - s_L}{b_H - s_L}$. Since $h > 1$, ex post efficiency cannot be achieved if the buyer is ex ante sufficiently likely to be high type. This is consistent with the argument preceding Proposition 1.

4.1 Second-best outcome in the seller-offer bargaining game

First, consider the seller-offer bargaining game. Assume $\pi_B > \frac{b_L - s_L}{b_H - s_L}$ so that ex post efficiency cannot be achieved in any acceptable CE. As discussed in the proof of Lemma 2, for any $\mu^{\text{SO}*}$, the ex ante expected social surplus is given by

$$\pi_S \pi_B (b_H - s_H) + (1 - \pi_S)\pi_B(b_H - s_L) + (1 - \pi_S)(1 - \pi_B)(1 - x^{\text{SO}}_{\text{LL}})(b_L - s_L)$$

Therefore, the acceptable CE that maximizes the expected social surplus can be computed by solving the following linear program in $x^{\text{SO}} = (x^{\text{SO}}_{\text{LH}}, x^{\text{SO}}_{\text{LL}})$:

$$\min_{x^{\text{SO}} \in [0,1]^2} x^{\text{SO}}_{\text{LL}}$$

$$\text{subject to} \quad (\text{SO-IC}_{s_L}) \text{ and } (\text{SO-IC}_{b_H}).$$

[26]The superscript "Eff" stands for ex post efficiency.

It is straightforward to verify that the constraint set contains only $x^{\text{SO}} = (1,1)$, which is then automatically the solution to the linear program. Hence, in the SB acceptable CE, the seller always offers the price b_H whenever there is a gain from trade. This leads to the following proposition.

Proposition 2. *If $\pi_B > \frac{b_L - s_L}{b_H - s_L}$, then, in the unique acceptable CE of the seller-offer bargaining game, the trader pairs (s_H, b_H) and (s_L, b_H) trade with probability one, while the pair (s_L, b_L) never trades.*

Proof. See Appendix C. □

As discussed in Section 3.1, in any acceptable CE, the high-type seller never offers the low price b_L, since doing so would yield him a negative payoff. If $\pi_B > \frac{b_L - s_L}{b_H - s_L}$, even the low-type seller cannot be incentivized to offer b_L because the buyer is sufficiently likely to be high type; he would prefer to offer the high price b_H, expecting the high-type buyer to accept it.

4.2 Second-best outcome in the mediated bargaining game

Next, consider the mediated bargaining game. Assume $h > \bar{h}^{\text{Eff}}$ so that ex post efficiency cannot be achieved in any acceptable CE. For any $\mu^{\text{MB}*}$, the ex ante expected social surplus is given by [27]

$$\begin{aligned} &\sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b) \sum_{q\in Q} \mu^{\text{MB}*}(q)(b-s)\cdot \mathbf{1}_{\{p_S(s)\le q(s,b)\le p_B(b)\}} \\ &= \pi_S\pi_B(b_H - s_H)x^{\text{MB}}_{\text{HH}} + (1-\pi_S)\pi_B(b_H - s_L) \\ &\quad + (1-\pi_S)(1-\pi_B)(b_L - s_L)(1 - x^{\text{MB}}_{\text{LL}}). \end{aligned} \tag{4.1}$$

The associated linear program is

$$\begin{aligned} &\max_{x^{\text{MB}}\in[0,1]^3} \quad (4.1) \\ &\text{subject to} \quad (\text{MB-IC}_{s_H}),\ (\text{MB-IC}_{s_L}),\ (\text{MB-IC}_{b_H}),\ (\text{IC}_{\text{Int}}\text{-}1),\ \text{and}\ (\text{IC}_{\text{Int}}\text{-}2). \end{aligned} \tag{MB-P}$$

[27] Since the trading price and commissions cancel out, the realized social surplus from a trade between types (s,b) is simply $b - s$.

By construction, trade always occurs for the pair (s_L, b_H). Thus, the trade-off lies between the expected social surplus from the high-type pair (s_H, b_H), given by $\pi_S \pi_B (b_H - s_H) x_{\text{HH}}^{\text{MB}}$, and that from the low-type pair (s_L, b_L), given by $(1 - \pi_S)(1 - \pi_B)(b_L - s_L)(1 - x_{\text{LL}}^{\text{MB}})$.

First, I show that (MB-IC$_{b_H}$) and (MB-IC$_{s_L}$) must bind in the SB outcome.

Lemma 5. *If $h > \bar{h}^{\text{Eff}}$, then* (MB-IC$_{b_H}$) *and* (MB-IC$_{s_L}$) *must bind at the solution to* (MB-P).

Proof. See Appendix D □

Recall that, on the equilibrium path, the high-type buyer obtains a positive payoff only when he trades at the price $p_B(b_L)$. (MB-IC$_{b_H}$) requires that this occurs more frequently when he truthfully reports his type than when he misreports; that is, $1 - x_{\text{LH}} \geq 1 - x_{\text{LL}}$.[28] If (MB-IC$_{b_H}$) does not bind, then I can increase $1 - x_{\text{LL}}$ without violating any constraint, thereby increasing the expected social surplus. This change is innocuous to the seller's incentive; the high-type seller is unaffected, and the low-type seller finds truthful reporting more attractive because his expected trade probability increases, while the payoffs from deviations remain unchanged. This change is also innocuous to the intermediary's incentive because she believes that the offer $p_B(b_L)$ is less likely to be accepted, making deviation to $p_B(b_L)$ when recommended $p_B(b_H)$ less attractive. A symmetric argument applies to the recommendation $p_B(b_L)$. Hence, I must have $x_{\text{LH}} = x_{\text{LL}}$ at the optimum.

Given that (MB-IC$_{b_H}$) binds, similar logic applies to (MB-IC$_{s_L}$). Specifically, I can either decrease x_{LH} and x_{LL}, or increase x_{HH} to raise the expected social surplus. As in the previous case, these adjustments do not violate the intermediary's IC constraints, since they affect her beliefs in ways that make deviations less attractive or obedience more attractive.

Define $y_0 = \frac{(1-\pi_B)(b_L - h s_L)}{\pi_B (b_H - b_L)}$, and let $y(x_{\text{L}})$ denote the value of x_{HH} that makes (MB-IC$_{s_L}$) bind, given that $x_{\text{LH}} = x_{\text{LL}} = x_{\text{L}}$:

$$y(x_{\text{L}}) = y_0 + (1 - y_0) x_{\text{L}}.$$

Then, by Lemma 5, when $h > \bar{h}^{\text{Eff}}$, the solution to (MB-P) must take the form $(y(x_{\text{L}}), x_{\text{L}}, x_{\text{L}})$.

[28] For notational simplicity, I omit the superscript "MB" hereafter.

Substituting this into (MB-P), I rewrite the program as follows:

$$\begin{aligned}\max_{x_{\mathrm{L}}\in[0,1]} \quad & \pi_S\pi_B(b_H - s_H)[y_0 + (1-y_0)x_{\mathrm{L}}] + (1-\pi_S)\pi_B(b_H - s_L) \\ & + (1-\pi_S)(1-\pi_B)(b_L - s_L)(1-x_{\mathrm{L}}) \\ \text{subject to} \quad & (\pi_S\pi_B b_H y_0 - J)x_{\mathrm{L}} \le \pi_S\pi_B b_H y_0 - \max\{J, 0\}, \end{aligned} \tag{4.2}$$

where (4.2) summarizes the remaining constraints, ($\mathrm{IC_{Int}}$-1) and ($\mathrm{IC_{Int}}$-2).[29] Since this is a linear program in a single variable with a single binding constraint, the key is to identify the thresholds that determine the sign of the objective's slope and the binding constraint.

First, observe that there exists a threshold h_{obj} such that the sign of the objective's slope changes at $h = h_{\mathrm{obj}}$.[30] To see this, note that increasing x_{L} makes truthful reporting more attractive for the low-type seller, as it increases his chance of obtaining the high price $p_B(b_H)$.[31] To maintain incentive compatibility, I must also make misreporting more attractive by increasing the probability x_{HH} that he obtains the high price when he misreports, according to $y(x_{\mathrm{L}})$. This creates a trade-off: increasing x_{L} reduces the trade probability for the low-type pair (s_L, b_L) but allows for a higher trade probability x_{HH} for the high-type pair (s_H, b_H). However, when h is small, the gain from increasing x_{L} is limited, as the low-type seller suffers a substantial loss from a higher probability of no trade with the low-type buyer. As a result, increasing x_{L} leads to a slight increase in x_{HH}, resulting in a net decline in expected social surplus.

Next, note that there exists a threshold h_J on the ratio h below which the coefficient for x_{L} in (4.2) is nonnegative:

$$J \le \pi_S\pi_B b_H y_0 \iff h \le \frac{b_L}{s_L} - \frac{(b_H - b_L)J}{\pi_S(1-\pi_B)b_H s_L} \equiv h_J.$$

[29] ($\mathrm{MB\text{-}IC}_{s_H}$) is reduced to $y_0(1 - x_{\mathrm{L}}) \ge 0$, which holds for any $x_{\mathrm{L}} \in [0, 1]$ since $y_0 \ge 0$.

[30] The expression for h_{obj} is

$$h_{\mathrm{obj}} = \frac{b_L}{s_L} - \frac{(b_H - b_L)(J - J_s)}{\pi_S(1-\pi_B)(b_H - s_H)s_L},$$

where

$$J_s = \pi_S\pi_B s_H + (1-\pi_S)\pi_B(b_H - b_L) - (1-\pi_S)(1-\pi_B)s_L.$$

[31] This effect dominates the effect of the reduction in trade probability with the low-type buyer if $\pi_B > \frac{b_L - hs_L}{b_H - hs_L} \Leftrightarrow h > \bar{h}^{\mathrm{Eff}}$.

Since the sign of J plays a critical role, as shown below, note also that

$$J \geq 0 \iff \pi_B \geq \frac{(1-\pi_S)b_L}{b_H} \equiv \pi_B(\pi_S \mid J=0).$$

Now suppose $h > h_J$. This case can arise only when $J > 0$, since $J < 0$ implies $h_J > \frac{b_L}{s_L} \geq \bar{h}$. Then, (4.2) forces $x_{\mathrm{L}} = 1$. Intuitively, when h is large, the intermediary's gain from commissions is too large to incentivize her to offer the low price $p_B(b_L)$. Thus, only $x_{\mathrm{L}} = 1$ is feasible. Since $(y(1), 1, 1) = (1, 1, 1)$ is the only feasible mediation plan, it is the solution to (MB-P). At this solution, the intermediary always offers the high price $p_B(b_H)$ whenever there is a gain from trade, so the high-type pair (s_H, b_H) always trades, while the low-type pair (s_L, b_L) never trades.

Next, consider $h \leq h_J$ and $J \geq 0$. Then, any $x_{\mathrm{L}} \in [0, 1]$ is feasible under (4.2). Although $J \geq 0$ implies that it is harder to incentivize the intermediary to offer $p_B(b_L)$ than $p_B(b_H)$, I can alter her belief by adjusting x_{HH} according to $y(x_{\mathrm{L}})$ so that the deviation to $p_B(b_H)$ becomes sufficiently unattractive. Hence, it is possible to incentivize the intermediary to offer $p_B(b_L)$ for any x_{L}. In contrast, if $J < 0$, then (4.2) forces $x_{\mathrm{L}} < 1$. In this case, it is harder to incentivize the intermediary to offer $p_B(b_H)$, and she would deviate to $p_B(b_L)$ if her belief remained at the prior. To make the recommendation $p_B(b_H)$ incentive compatible, it is thus necessary to reduce $x_{\mathrm{L}} < 1$, thereby strengthening her belief that $p_B(b_L)$ would be rejected.

By the above argument, if $h \leq \min\{h_J, h_{\mathrm{obj}}\} \equiv h^*$, then $x_{\mathrm{L}} = 0$ is both feasible and optimal, as the objective is decreasing in x_{L}. The SB outcome is characterized by $(y(0), 0, 0) = (y_0, 0, 0)$, where the intermediary offers the high price $p_B(b_H)$ to the high-type pair with probability less than one, and always offers the low price $p_B(b_L)$ to both (s_L, b_H) and (s_L, b_L). Accordingly, the high-type pair trades with probability less than one, while the low-type pair always trades.

If $h_{\mathrm{obj}} < h \leq h_J$, then the objective is increasing in x_{L}, so the optimal x_{L} lies at the right endpoint of the feasible set. If $\pi_B \geq \pi_B(\pi_S \mid J = 0)$, then the solution is $(y(1), 1, 1) = (1, 1, 1)$. If $\pi_B < \pi_B(\pi_S \mid J = 0)$, the solution entails $x_{\mathrm{L}} < 1$. The former case is the same as the case when $h > h_J$. In the latter case, whenever there is a gain from trade, the intermediary offers both the low price $p_B(b_L)$ and the high price $p_B(b_H)$ with positive probability. As a result, both the high-type pair and the low-type pair trade with probability less than one.

The above analysis leads to the following proposition.[32] For expositional clarity, I focus on the case where $\frac{b_H}{s_H} > \frac{b_L}{s_L}$ so that $\bar{h} = \frac{b_L}{s_L}$.[33]

Proposition 3. *If $h > \bar{h}^{\text{Eff}}$ and $\frac{b_H}{s_H} > \frac{b_L}{s_L}$, then in the SB outcome:*

1. *If $h \leq h^*$, the trader pair (s_H, b_H) trades with probability less than one, while the pair (s_L, b_L) trades with probability one. The associated ex ante expected social surplus is decreasing in h.*

2a. *If $h > h_J$ or $h_{\text{obj}} < h \leq h_J$, and $\pi_B \geq \pi_B(\pi_S \mid J = 0)$, the pair (s_H, b_H) trades with probability one, while the pair (s_L, b_L) never trades. The associated ex ante expected social surplus is constant in h.*

2b. *If $h_{\text{obj}} < h < h_J$ and $\pi_B < \pi_B(\pi_S \mid J = 0)$, both (s_H, b_H) and (s_L, b_L) trade with probability less than one. The associated ex ante expected social surplus is decreasing in h.*

Proof. See Appendix E. □

As a result, for any given parameter configuration, the SB level of the ex ante expected social surplus is weakly decreasing in the ratio h. This indicates that higher commission costs may reduce the efficiency of the bargaining outcome.

Note that Proposition 3 does not rule out the case where $h^* > \bar{h}$ or $h^* \leq 1$. In the former case, h^* does not bind, and the SB outcome corresponds to Case 1 for all $h \leq \bar{h}$. In the latter case, h^* does not bind either, and the SB outcome corresponds to Case 2a for all $h \leq \bar{h}$, since $h^* \leq 1$ is incompatible with $\pi_B < \pi_B(\pi_S \mid J = 0)$. If instead $\frac{b_H}{s_H} \leq \frac{b_L}{s_L}$, then both Case 1 and Case 2a can arise, but Case 2b does not.[34]

[32]In the statement of the proposition, I omit the trade probability for the trader pair (s_L, b_H) because they always trade by construction. I also omit the knife-edge case where $h = h_{\text{obj}}$, since in this case all mediation plans of the form $(y(x_{\text{L}}), x_{\text{L}}, x_{\text{L}})$ yield the same expected social surplus.

[33]Note that $\frac{b_H}{s_H} > \frac{b_L}{s_L}$ implies $b_H - s_H > b_L - s_L$ because

$$\frac{b_H}{s_H} > \frac{b_L}{s_L} \iff (b_H - s_H)b_L > b_H(b_L - s_L) \Longrightarrow b_H - s_H > b_L - s_L.$$

This implies that the social surplus generated by the high-type pair (s_H, b_H) is higher than that generated by the low-type pair (s_L, b_L):

[34]This is because $\frac{b_H}{s_H} \leq \frac{b_L}{s_L}$ implies $h_J < h_{\text{obj}}$ for all π_B. See Appendix F for a complete characterization.

Finally, I compare the expected social surplus in the seller-offer and the mediated bargaining games. Propositions 2 and 3 imply that the SB levels of expected social surplus coincide under the two games if Case 2a arises in the latter. Although the mediated bargaining game may yield a strictly lower expected social surplus when Case 2b arises, in all other cases it yields a weakly higher expected social surplus, regardless of the value of h.[35] It is worth noting that when $h \leq h^*$, the mediated bargaining game strictly outperforms the seller-offer bargaining game in terms of expected social surplus. This result is summarized in the following corollary:

Corollary 1. *If $h \leq h^*$, the mediated bargaining game can achieve a higher ex ante expected social surplus than the seller-offer bargaining game in the second-best scenario.*

Thus, even when the intermediary is biased, her mediation can improve the efficiency of the bargaining outcome. This provides a rationale for the widespread use of intermediaries in bargaining, even when their bias is common knowledge.

5 Conclusion

This paper has examined how a biased intermediary who lacks both commitment and enforcement power can affect bargaining outcomes. To this end, I considered a minimal departure from a seller-offer bargaining game by introducing a weak intermediary—one who aligns with the seller's interests, possesses the same instruments, and has no private information or expertise. The main result shows that even such a weak intermediary can improve the efficiency of the bargaining outcome, providing a rationale for the widespread use of intermediaries, even when their bias is common knowledge.

The main result generalizes to other payoff specifications. What is essential is that the intermediary strictly prefers trade to no trade and that her payoff increases with the price. Combined with her lack of commitment, this creates an incentive to offer only the buyer's maximum acceptable prices. For example, consider the following specification: if the traders' types are (s, b) and they trade at a price p, then the seller obtains $p - s$, the buyer obtains $b - p$,

[35] Note that Case 2a is associated with $x_L = 1$. In Case 2b, although the expected social surplus is increasing in x_L, the intermediary's IC constraint (4.2) restricts x_L to be less than one. Therefore, the expected social surplus is higher in Case 2a than in Case 2b.

and the intermediary obtains $\alpha(p-s)+(1-\alpha)(b-p)$, where $\alpha \in [0,1]$. If $\alpha \in \left(\frac{1}{2}, 1\right]$, I can obtain qualitatively the same results. This specification aligns with the setting of Loertscher and Marx (2022), who study bilateral bargaining and model the market as a mechanism that maximizes the expected weighted welfare of the firms (sellers and buyers). In this sense, the intermediary can be interpreted as an explicit agent implementing part of the market mechanism. Alternatively, suppose that the traders' payoffs remain as above, but the intermediary obtains $\alpha(p-s)$ if trade occurs and $-c$ otherwise, where $c \in \mathbb{R}_+$. Then, as long as $\alpha > 0$, qualitatively the same results hold. This setup aligns with that of Kydd (2003), who studies a biased intermediary in the context of international relations.

An interesting extension would be to move beyond binary types while maintaining the private-value assumption. Although the binary structure plays a role in the current analysis, the key forces remain in more general settings—namely, that the player making the offer is incentivized to propose only the buyer's maximum acceptable prices, and that the seller never offers a price below his own valuation.

Another promising direction is to endogenize the commission structure. Throughout the paper, the commission rates were assumed to be exogenously fixed. As discussed in footnote 2, in the U.S. real estate industry, the seller has traditionally paid about 5% of the sale price as commission, which is then split equally between the seller's and the buyer's agents. However, this practice has recently changed, allowing the seller to choose not to pay the buyer's agent, leaving the buyer to negotiate directly with their own agent about the commission (see Kamin, 2024). In light of this reform, allowing the traders and the intermediary to bargain over the commission structure would be a natural extension and could contribute to the ongoing policy debate. I plan to explore these extensions in future research.

Appendix

This appendix is divided into several sections. Section A provides the proof of Lemma 3. Section B provides the proof of Lemma 4. Section C provides the proof of Proposition 2. Section E provides the proof of Proposition 3. Section F provides a complete characterization of the second-best outcome in the mediated bargaining game.

A Proof of Lemma 3

I prove the contraposition. Consider a mediation plan $\mu^{\text{MB}} \in \Delta\left(Q \times \left\{r^{\text{MB}*}\right\}\right)$ that violates (3.4) for some $q \in \operatorname{supp}\left(\mu^{\text{MB}}\right)$. I show that the intermediary has a profitable manipulation. Suppose that the intermediary receives a recommendation $p \notin \{p_B(b_L), p_B(b_H)\}$. Her posterior belief that the traders' types are (s, b) is

$$\nu^{\text{MB}}(s, b \mid p) = \frac{\Pr(s, b) \sum_{q:\, q(s,b)=p} \mu^{\text{MB}}(q)}{\sum_{(\tilde{s},\tilde{b}) \in \Theta_S \times \Theta_B} \Pr(\tilde{s}, \tilde{b}) \sum_{q:\, q(\tilde{s},\tilde{b})=p} \mu^{\text{MB}}(q)}.$$

By assumption, at least one of $\nu^{\text{MB}}(s_H, b_H \mid p)$, $\nu^{\text{MB}}(s_L, b_H \mid p)$, or $\nu^{\text{MB}}(s_L, b_L \mid p)$ is positive. Assuming that the traders are honest and obedient, if the intermediary offers a price $\tilde{p}$, her expected payoff is

$$\begin{cases} \left[\nu^{\text{MB}}(s_L, b_H \mid p) + \nu^{\text{MB}}(s_L, b_L \mid p)\right](\delta_S + \delta_B)\tilde{p} & \text{if } \tilde{p} \in [p_S(s_L), p_B(b_L)]; \\ \nu^{\text{MB}}(s_L, b_H \mid p)(\delta_S + \delta_B)\tilde{p} & \text{if } \tilde{p} \in (p_B(b_L), p_S(s_H)); \\ \left[\nu^{\text{MB}}(s_H, b_H \mid p) + \nu^{\text{MB}}(s_L, b_H \mid p)\right](\delta_S + \delta_B)\tilde{p} & \text{if } \tilde{p} \in [p_S(s_H), p_B(b_H)]; \\ 0 & \text{if } \tilde{p} \notin [p_S(s_L), p_B(b_H)]. \end{cases}$$

In any case, she can gain by deviating to either $p_B(b_L)$ or $p_B(b_H)$. □

B Proof of Lemma 4

Given Lemma 3, consider an arbitrary mediation plan $\mu^{\text{MB}} \in \Delta\left(Q \times \left\{r^{\text{MB}*}\right\}\right)$ that satisfies (3.4) for all $q \in \operatorname{supp}\left(\mu^{\text{MB}}\right)$. By construction, all players obtain the same expected payoffs under μ^{MB}

and $\mu^{\text{MB}*}$ if they are honest and obedient. Hence, it suffices to show that μ^{MB} allows more room for profitable manipulation by all players than $\mu^{\text{MB}*}$.

As in the seller-offer bargaining game, let $x_{\text{HH}}^{\text{MB}}$ and $x_{\text{LL}}^{\text{MB}}$ denote the total probability under μ^{MB} that the price $p_B(b_H)$ is recommended for the report (s_H, b_H) and (s_L, b_L), respectively:

$$\begin{aligned} x_{\text{HH}}^{\text{MB}} &= \sum_{q:\, q(s_H,b_H)=p_B(b_H)} \mu^{\text{MB}}(q), \\ x_{\text{LL}}^{\text{MB}} &= \sum_{q:\, q(s_L,b_L)=p_B(b_H)} \mu^{\text{MB}}(q). \end{aligned}$$

Note that these probabilities coincide under μ^{MB} and $\mu^{\text{MB}*}$ because the recommendations for (s_H, b_H) and (s_L, b_L) remain unchanged.

As the difference between μ^{MB} and $\mu^{\text{MB}*}$ does not affect the IC constraints of the high-type seller, the low-type buyer, and the intermediary, it remains to show that those of the low-type seller and the high-type buyer are (weakly) more stringent under μ^{MB} than under $\mu^{\text{MB}*}$.[36]

Low-type seller. If the low-type seller misreports his type under $\mu^{\text{MB}*}$, he can obtain a positive payoff only when the buyer is high type; he can obatin $(1-\delta_S)p_B(b_H) - s_L = \frac{b_H}{h} - s_L$ if the price $p_B(b_H)$ is offered and $(1-\delta_S)p_B(b_L) - s_L = \frac{b_L}{h} - s_L$ if $p_B(b_L)$ is offered. Hence, his expected payoff from misreporting under $\mu^{\text{MB}*}$ is at most

$$\pi_B \left[x_{\text{HH}}^{\text{MB}} \frac{b_H}{h} + \left(1 - x_{\text{HH}}^{\text{MB}}\right) \frac{b_L}{h} - s_L \right].$$

Under μ^{MB}, he may additionally obtain a positive payoff when the buyer is low type and the intermediary offers a price $p \in (p_S(s_L), p_B(b_L)]$. Thus, his expected payoff from misreporting

[36]This is because, for the high-type seller and the low-type buyer, the price $q(s_H, b_L)$ appears in their expected payoff only when they report their type truthfully, in which case they cannot do better than following $r^{\text{MB}*}$. For the intermediary, if she receives the recommendation $q(s_H, b_L)$, she learns that trade never occurs for this trader pair, and hence following the recommendation is always optimal.

under μ^{MB} is at most

$$\pi_B \left[x_{\text{HH}}^{\text{MB}} \frac{b_H}{h} + \left(1 - x_{\text{HH}}^{\text{MB}}\right) \frac{b_L}{h} - s_L \right] + (1 - \pi_B) \sum_{p \in (p_S(s_L), p_B(b_L)]} \sum_{q:\, q(s_H, b_L) = p} \mu^{\text{MB}}(q)[(1 - \delta_S)p - s_L],$$

where the second term is nonnegative. Therefore, his IC constraint is weakly more stringent under μ^{MB} than under $\mu^{\text{MB}*}$.

High-type buyer. If the high-type buyer misreports his type under $\mu^{\text{MB}*}$, he can obtain a payoff of $b_H - (1 + \delta_B)p_B(b_L) = b_H - b_L$ when the seller is low type and the price $p_B(b_L)$ is offered. In all other cases, he cannot obtain a positive payoff. Hence, his expected payoff from misreporting under $\mu^{\text{MB}*}$ is at most

$$(1 - \pi_S)(1 - x_{\text{LL}}^{\text{MB}})(b_H - b_L).$$

Under μ^{MB}, he may additionally obtain a positive payoff when the seller is high type and the intermediary offers a price $p \in [p_S(s_H), p_B(b_H))$. Thus, his expected payoff from misreporting under μ^{MB} is at most

$$(1 - \pi_S)(1 - x_{\text{LL}}^{\text{MB}})(b_H - b_L) + \pi_S \sum_{p \in [p_S(s_H), p_B(b_H))} \sum_{q:\, q(s_H, b_L) = p} \mu^{\text{MB}}(q)[b_H - (1 + \delta_B)p],$$

where the second term is nonnegative. Therefore, his IC constraint is also weakly more stringent under μ^{MB} than under $\mu^{\text{MB}*}$.

Finally, consider the ex ante expected social surplus. Since the trading price and commissions cancel out, the realized social surplus from a trade between types (s, b) is simply $b - s$. Hence, the ex ante expected social surplus under μ^{MB} is given by

$$\begin{aligned}&\sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b) \sum_{q\in Q} \mu^{\text{MB}}(q)(b-s)\cdot \mathbf{1}_{\{p_S(s)\leq q(s,b)\leq p_B(b)\}} \\ &= \pi_S\pi_B x^{\text{MB}}_{\text{HH}}(b_H - s_H) + (1-\pi_S)\pi_B(b_H - s_L) \\ &\quad + (1-\pi_S)(1-\pi_B)(1-x^{\text{MB}}_{\text{LL}})(b_L - s_L),\end{aligned}$$

which depends only on $x^{\text{MB}}_{\text{HH}}$ and $x^{\text{MB}}_{\text{LL}}$, implying that the expected social surplus coincides under the two mediation plans. □

C Proof of Proposition 2

I show that only $x^{\text{SO}} = (1,1)$ satisfies the IC constraints. When $\pi_B > \frac{b_L - s_L}{b_H - s_L}$, (SO-IC$_{s_L}$) is equivalent to (SO-IC$_{s_L}$-2), which can be rewritten as

$$x^{\text{SO}}_{\text{LL}} \leq 1 - \frac{\pi_B(b_H - b_L)}{(1-\pi_B)(b_L - s_L)}\left(1 - x^{\text{SO}}_{\text{LH}}\right). \tag{C.1}$$

In $x^{\text{SO}}_{\text{LH}}$–$x^{\text{SO}}_{\text{LL}}$ space, (C.1) defines the region below a linear function that passes through $(1,1)$ and has a slope greater than one. At the same time, (SO-IC$_{b_H}$) defines the region above the 45-degree line. Thus, the only point satisfying both constraints is $x^{\text{SO}} = (1,1)$. □

D Proof of Lemma 5

(MB-IC$_{b_H}$) must bind

Suppose that (MB-IC$_{b_H}$) does not bind at the solution $x = (x_{\text{HH}}, x_{\text{L}}, x_{\text{L}})$; that is,

$$x_{\text{LL}} > x_{\text{LH}}.$$

Then, I can slightly decrease x_{LL} without violating any constraint. Since decreasing x_{LL} increases the expected social surplus, this contradicts the optimality of x. Therefore,

$$x_{\mathrm{LH}} = x_{\mathrm{LL}} \equiv x_{\mathrm{L}}$$

must hold at the optimum.

In this case, (MB-IC$_{s_H}$) is wrtitten as $x_{\mathrm{HH}} \geq x_{\mathrm{L}}$, and the low-type seller's IC constraint (MB-IC$_{s_L}$) and the intermediary's IC constraints (IC$_{\mathrm{Int}}$-1) and (IC$_{\mathrm{Int}}$-2) are respectively rewritten as follows:

$$[\pi_B(b_H - hs_L) - b_L + hs_L]x_{\mathrm{L}} + (1-\pi_B)(b_L - hs_L) \geq \pi_B(b_H - b_L)x_{\mathrm{HH}}, \tag{D.1}$$

$$\pi_S\pi_B b_H x_{\mathrm{HH}} \geq (1-\pi_S)[(1-\pi_B)b_L - \pi_B(b_H - b_L)]x_{\mathrm{L}}, \tag{D.2}$$

$$(1-\pi_S)[(1-\pi_B)b_L - \pi_B(b_H - b_L)](1-x_{\mathrm{L}}) \geq \pi_S\pi_B b_H(1-x_{\mathrm{HH}}). \tag{D.3}$$

(MB-IC$_{s_L}$) must bind

I now show that (D.1) must bind at the optimum. First, suppose that $\pi_B > \frac{b_L}{b_H}$. In this case, the left-hand side of (D.3) is at most zero, while the right-hand side is nonnegative. Hence, (D.3) can be satisfied only when $x_{\mathrm{HH}} = x_{\mathrm{L}} = 1$. Moreover, if $x_{\mathrm{L}} = 1$ at the optimum, then $x_{\mathrm{HH}} = 1$ necessarily hold by (MB-IC$_{s_H}$), and hence (D.1) also binds. Therefore, if $\pi_B > \frac{b_L}{b_H}$ or $x_{\mathrm{L}} = 1$ holds at the optimmum, then (D.1) must bind.

Now suppose instead that $\pi_B \leq \frac{b_L}{b_H}$ and that the solution satisfies $x_{\mathrm{L}} < 1$ and

$$[\pi_B(b_H - hs_L) - b_L + hs_L]x_{\mathrm{L}} + (1-\pi_B)(b_L - hs_L) > \pi_B(b_H - b_L)x_{\mathrm{HH}}. \tag{D.4}$$

Case 1: $x_{\mathrm{L}} > 0$

Since $\pi_B \leq \frac{b_L}{b_H}$, the coefficient of x_{L} in (D.2) (resp. the coefficient of $1 - x_{\mathrm{L}}$ in (D.3)) is nonnegative. Thus, I can slightly decrease x_{L} without violating any constraint. Since decreasing x_{L} increases the expected social surplus, this contradicts the optimality of x.

Case 2: $x_{\mathrm{L}} = 0$

In this case, (D.4) simplifies to

$$x_{\mathrm{HH}} < \frac{(1-\pi_B)(b_L - hs_L)}{\pi_B(b_H - b_L)} \equiv y_0.$$

As $h > \bar{h}^{\mathrm{Eff}}$ is equivalent to $y_0 < 1$, I can slightly increase x_{HH} without violating any constraint. Since increasing x_{HH} raises the expected social surplus, this contradicts the optimality of x.

Thus, (MB-IC$_{s_L}$) must bind at the optimum.

Therefore, under $h > \bar{h}^{\mathrm{Eff}}$, the solution to (MB-P) must take the form $(y(x_{\mathrm{L}}), x_{\mathrm{L}}, x_{\mathrm{L}})$, where

$$y(x_{\mathrm{L}}) = y_0 + (1 - y_0)x_{\mathrm{L}}$$

is the value of x_{HH} that makes (MB-IC$_{s_L}$) bind, given that $x_{\mathrm{LH}} = x_{\mathrm{LL}} = x_{\mathrm{L}}$. □

E Proof of Proposition 3

Note that (4.2) is reduced to the following constraint on x_{L}:

- $x_{\mathrm{L}} \leq 1$ if $h \leq h_J$ and $J \geq 0$;
- $x_{\mathrm{L}} \geq 1$ if $h > h_J$ and $J \geq 0$:
- $x_{\mathrm{L}} \leq \frac{\pi_S \pi_B b_H y_0}{\pi_S \pi_B b_H y_0 - J} \equiv x_{\mathrm{L}}^*$ if $h < h_J$ and $J < 0$.

Case 1: $h \leq h^*$**.**

Since $x_{\mathrm{L}} = 0$ is feasible and the objective is decreasing in x_{L}, the solution to (MB-P) is

$$x = (y(0), 0, 0) = (y_0, 0, 0).$$

This mediation plan induces the trade pattern described in Case 1 of the proposition.

Case 2a: $h > h_J$ **or** $h_{\text{obj}} < h \leq h_J$**, and** $\pi_B \geq \pi_B(\pi_S \mid J = 0)$**.**

In this case, the objective is maximized at $x_{\text{L}} = 1$, either because (4.2) forces $x_{\text{L}} \geq 1$ (as $h > h_J$), or because any $x_{\text{L}} \in [0, 1]$ is feasible and the objective is increasing in x_{L} (as $h_{\text{obj}} < h \leq h_J$). In either case, the solution to (MB-P) is

$$x = (y(1), 1, 1) = (1, 1, 1).$$

This mediation plan induces the trade pattern described in Case 2a of the proposition.

Case 2b: $h_{\text{obj}} < h < h_J$ **and** $\pi_B < \pi_B(\pi_S \mid J = 0)$**.**

Since any $x_{\text{L}} \in \left[0, x_{\text{L}}^*\right]$ is feasible and the objective is increasing in x_{L}, the solution to (MB-P) is

$$x = \left(y\left(x_{\text{L}}^*\right), x_{\text{L}}^*, x_{\text{L}}^*\right).$$

Since both $y\left(x_{\text{L}}^*\right)$ and x_{L}^* are less than one, this mediation plan induces the trade pattern described in Case 2b of the proposition.

Next, I examine the expected social surplus in each case.

- In Case 1, the expected social surplus is an increasing function of y_0. Since y_0 is decreasing in h, the expected social surplus is also decreasing in h.
- In Case 2a, the expected social surplus is a constant and independent of h.
- In Case 2b, the expected social surplus is given by

$$\pi_S \pi_B (b_H - s_H) \frac{y_0(\pi_S \pi_B b_H - J)}{\pi_S \pi_B b_H y_0 - J} + (1 - \pi_S)\pi_B(b_H - s_L)$$
$$- (1 - \pi_S)(1 - \pi_B)(b_L - s_L)\frac{J}{\pi_S \pi_B b_H y_0 - J}.$$

  Note that the sign of its derivative with respect to y_0 is the same as

$$-\pi_S \pi_B J[(b_H - s_H)(\pi_S \pi_B b_H - J) - (1 - \pi_S)(1 - \pi_B) b_H (b_L - s_L)].$$

Since $J < 0$ in this case, the expected social surplus is increasing in y_0 and hence decreasing in h if the term inside the bracket is positive. That is,

$$(b_H - s_H)(\pi_S \pi_B b_H - J) - (1-\pi_S)(1-\pi_B) b_H (b_L - s_L) > 0$$
$$\iff \pi_B \leq \frac{b_H s_L - s_H b_L}{b_H(b_H - s_H - b_L + s_L)} \equiv \pi_B^\star.$$

Observe that Case 2b arises only when $h_{\text{obj}} < h_J$, which, under the assumption $\frac{b_H}{s_H} > \frac{b_L}{s_L}$, is equivalent to $\pi_B < \pi_B^\star$. This completes the proof. □

F Complete Characterization of the Second-Best Outcome

In this section, I provide a complete characterization of the SB outcome. Recall that the two thresholds h_J and h_{obj} are given by

$$h_J = \frac{b_L}{s_L} - \frac{(b_H - b_L)J}{\pi_S(1-\pi_B) b_H s_L},$$
$$h_{\text{obj}} = \frac{b_L}{s_L} - \frac{(b_H - b_L)(J - J_s)}{\pi_S(1-\pi_B)(b_H - s_H) s_L}.$$

Note also that $h_J > \bar{h}^{\text{Eff}}$ if $\pi_B < \frac{b_L}{b_H}$, and that $h_{\text{obj}} > \bar{h}^{\text{Eff}}$ for all π_B. Define the following thresholds on π_B:

$$J - J_s \geq 0 \iff \pi_B \geq \frac{(1-\pi_S)(b_L - s_L)}{\pi_S(b_H - s_H) + (1-\pi_S)(b_L - s_L)} \equiv \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L),$$
$$h_J > 1 \iff \pi_B < \frac{\pi_S b_H (b_L - s_L) + (1-\pi_S)(b_H - b_L) b_L}{b_H[\pi_S(b_H - s_L) + (1-\pi_S)(b_H - b_L)]} \equiv \pi_B(\pi_S \mid h_J = 1),$$
$$h_{\text{obj}} > 1 \iff \pi_B < \frac{(b_L - s_L)[\pi_S(b_H - s_H) + (1-\pi_S)(b_H - b_L)]}{\pi_S(b_H - s_H)(b_H - s_L) + (1-\pi_S)(b_H - b_L)(b_L - s_L)} \equiv \pi_B(\pi_S \mid h_{\text{obj}} = 1).$$

When $\frac{b_H}{s_H} > \frac{b_L}{s_L}$.

In this case, $\bar{h} = \frac{b_L}{s_L}$. Since $\frac{b_H}{s_H} > \frac{b_L}{s_L}$ implies $b_H - s_H > b_L - s_L$, it follows that $\pi_B(\pi_S \mid J = 0) \leq \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L) \Leftrightarrow \pi_S \leq \pi_S^\star$, where

$$\pi_S^\star = \frac{(b_H - b_L)(b_L - s_L)}{b_L(b_H - s_H - b_L + s_L)}.$$

Table F.1: Definition of each region in the π_S–π_B space

Region	**Condition on** (π_S, π_B)
1	$\max\{\pi_B^\star, \pi_B(\pi_S \mid J=0)\} \leq \pi_B \leq \pi_B(\pi_S \mid h_J = 1)$
1a	$\pi_B(\pi_S \mid h_J = b_H/s_H) < \pi_B \leq \pi_B(\pi_S \mid h_J = 1)$
1b	$\pi_B(\pi_S \mid J=0) \leq \pi_B \leq \pi_B(\pi_S \mid h_J = b_H/s_H)$
2	$\pi_B(\pi_S \mid J=0) \leq \pi_B \leq \min\{\pi_B^\star, \pi_B(\pi_S \mid h_{\text{obj}} = 1)\}$
3	$\pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L) \leq \pi_B < \pi_B(\pi_S \mid J=0)$
4	$\pi_B < \min\{\pi_B(\pi_S \mid J=0), \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L)\}$
5	$\pi_B > \min\{\pi_B(\pi_S \mid h_J = 1), \pi_B(\pi_S \mid h_{\text{obj}} = 1)\}$

Note: Each region is defined as the set of (π_S, π_B) that satisfies the corresponding inequality shown in the table. See Figures F.1 and F.2 for graphical illustrations.

In addition, $h_J \leq h_{\text{obj}} \Leftrightarrow \pi_B \geq \pi_B^\star$ and $\pi_S^\star, \pi_B^\star \in (0,1)$. Thus, there are five distinct regions in the π_S–π_B space, as illustrated in Figure F.1. Table F.1 summarizes their definitions. In each region, the solution is given as follows:

1. If (π_S, π_B) lies in Region 1, then $\bar{h}^{\text{Eff}} < 1 \leq h_J \leq \min\left\{h_{\text{obj}}, \frac{b_L}{s_L}\right\}$. Hence, the solution is $(y_0, 0, 0)$ if $h \leq h_J$, and $(1,1,1)$ otherwise.

2. If (π_S, π_B) lies in Region 2, then $\max\left\{1, \bar{h}^{\text{Eff}}\right\} \leq h_{\text{obj}} \leq h_J \leq \frac{b_L}{s_L}$ and $\pi_B \geq \pi_B(\pi_S \mid J = 0)$. Hence, the solution is $(y_0, 0, 0)$ if $h \leq h_{\text{obj}}$, and $(1,1,1)$ otherwise.

3. If (π_S, π_B) lies in Region 3, then $\max\left\{1, \bar{h}^{\text{Eff}}\right\} < h_{\text{obj}} \leq \frac{b_L}{s_L} \leq h_J$ and $\pi_B < \pi_B(\pi_S \mid J = 0)$. Hence, the solution is $(y_0, 0, 0)$ if $h \leq h_{\text{obj}}$, and $\left(y(x_{\text{L}}^*), x_{\text{L}}^*, x_{\text{L}}^*\right)$ otherwise.

4. If (π_S, π_B) lies in Region 4, then $h^* > \frac{b_L}{s_L}$. Hence, the solution is $(y_0, 0, 0)$ for all $h \leq \frac{b_L}{s_L}$.

5. If (π_S, π_B) lies in Region 5, then $h^* < 1$. Hence, the solution is $(1,1,1)$ for all $h \leq \frac{b_L}{s_L}$.

When $\frac{b_H}{s_H} \leq \frac{b_L}{s_L}$ **and** $b_H - s_H > b_L - s_L$**.**

In this case, $\pi_B(\pi_S \mid J = 0) \leq \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L) \Leftrightarrow \pi_S \leq \pi_S^\star$ and $h_J \leq h_{\text{obj}} \Leftrightarrow \pi_B \geq \pi_B^\star$. Since $\pi_S^\star \geq 1$ and $\pi_B^\star \leq 0$, it follows that $\pi_B(\pi_S \mid J = 0) < \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L)$ for all π_S and $h_J < h_{\text{obj}}$ for all $(\pi_S, \pi_B) \in (0,1)^2$. Given that $\bar{h} = \frac{b_H}{s_H}$ in this case, it is useful to consider the following threshold:

$$h_J < \frac{b_H}{s_H} \iff \pi_B > \frac{\pi_S b_H (s_H b_L - b_H s_L) + (1-\pi_S) s_H (b_H - b_L) b_L}{b_H[\pi_S b_H (s_H - s_L) + (1-\pi_S) s_H (b_H - b_L)]} \equiv \pi_B(\pi_S \mid h_J = b_H/s_H).$$

Thus, there are four distinct regions in the π_S–π_B space, as illustrated in Figure F.2. Table F.1 summarizes their definitions. In each region, the solution is given as follows:

1a. If (π_S, π_B) lies in Region 1a, then $\max\left\{1, \bar{h}^{\text{Eff}}\right\} \leq h_J < \min\left\{h_{\text{obj}}, \frac{b_H}{s_H}\right\}$. Hence, the solution is $(y_0, 0, 0)$ if $h \leq h_J$, and $(1, 1, 1)$ otherwise.

1b. If (π_S, π_B) lies in Region 1b, then $h_J \geq \frac{b_H}{s_H}$. Hence, the solution is $(y_0, 0, 0)$ for all $h \leq \frac{b_H}{s_H}$.

4. If (π_S, π_B) lies in Region 4, then $h_J > \frac{b_L}{s_L} \geq \frac{b_H}{s_H}$. Hence, the solution is $(y_0, 0, 0)$ for all $h \leq \frac{b_H}{s_H}$.

5. If (π_S, π_B) lies in Region 5, then $h_J < 1$. Hence, the solution is $(1, 1, 1)$ for all $h \leq \frac{b_H}{s_H}$.

When $\frac{b_H}{s_H} \leq \frac{b_L}{s_L}$ and $b_H - s_H \leq b_L - s_L$.

When $b_H - s_H < b_L - s_L$, then $\pi_B(\pi_S \mid J = 0) \leq \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L) \Leftrightarrow \pi_S \geq \pi_S^\star$ and $h_J \leq h_{\text{obj}} \Leftrightarrow \pi_B \leq \pi_B^\star$. Since $\pi_S^\star < 0$ and $\pi_B^\star > 1$, it follows that $\pi_B(\pi_S \mid J = 0) < \pi_B(\pi_S \mid h_{\text{obj}} = b_L/s_L)$ for all π_S and $h_J < h_{\text{obj}}$ for all $(\pi_S, \pi_B) \in (0, 1)^2$. The same conclusion holds when $b_H - s_H = b_L - s_L$. Therefore, this case is analytically equivalent to the previous one, and I omit further discussion.

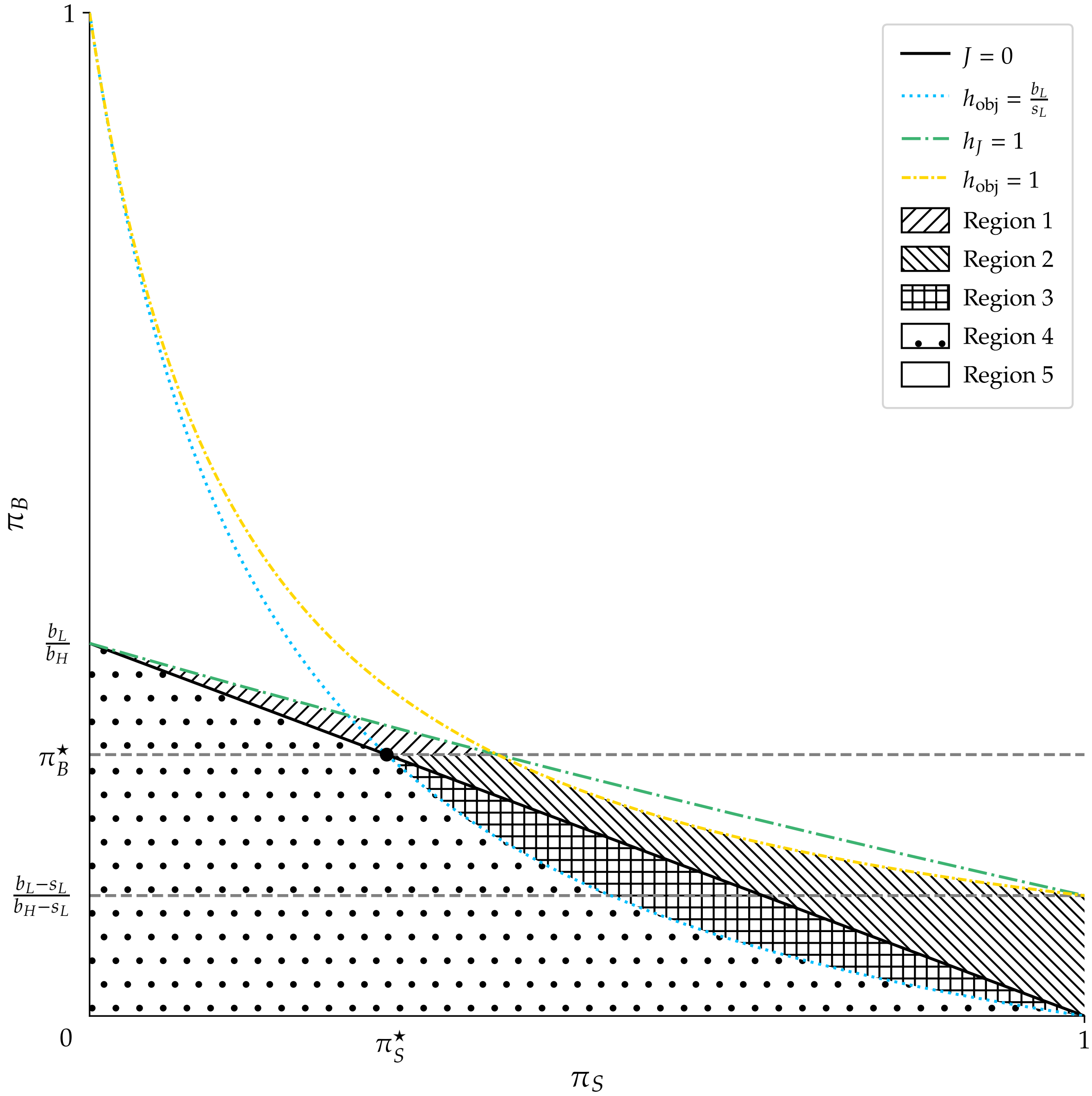

Figure F.1: Five distinct regions in the π_S–π_B space when $\frac{b_H}{s_H} > \frac{b_L}{s_L}$. The valuations are set at $b_H = 3.5$, $s_H = 1.5$, $b_L = 1.3$, and $s_L = 1.0$. The black solid, blue dotted, green dash–dot, and yellow dash–dot lines correspond to $\pi_B(\cdot \mid J = 0)$, $\pi_B(\cdot \mid h_{\text{obj}} = b_L/s_L)$, $\pi_B(\cdot \mid h_J = 1)$, and $\pi_B(\cdot \mid h_{\text{obj}} = 1)$, respectively. Each region is defined as the set of $(\pi_S, \pi_B) \in (0, 1)^2$ satisfying the following constraint:

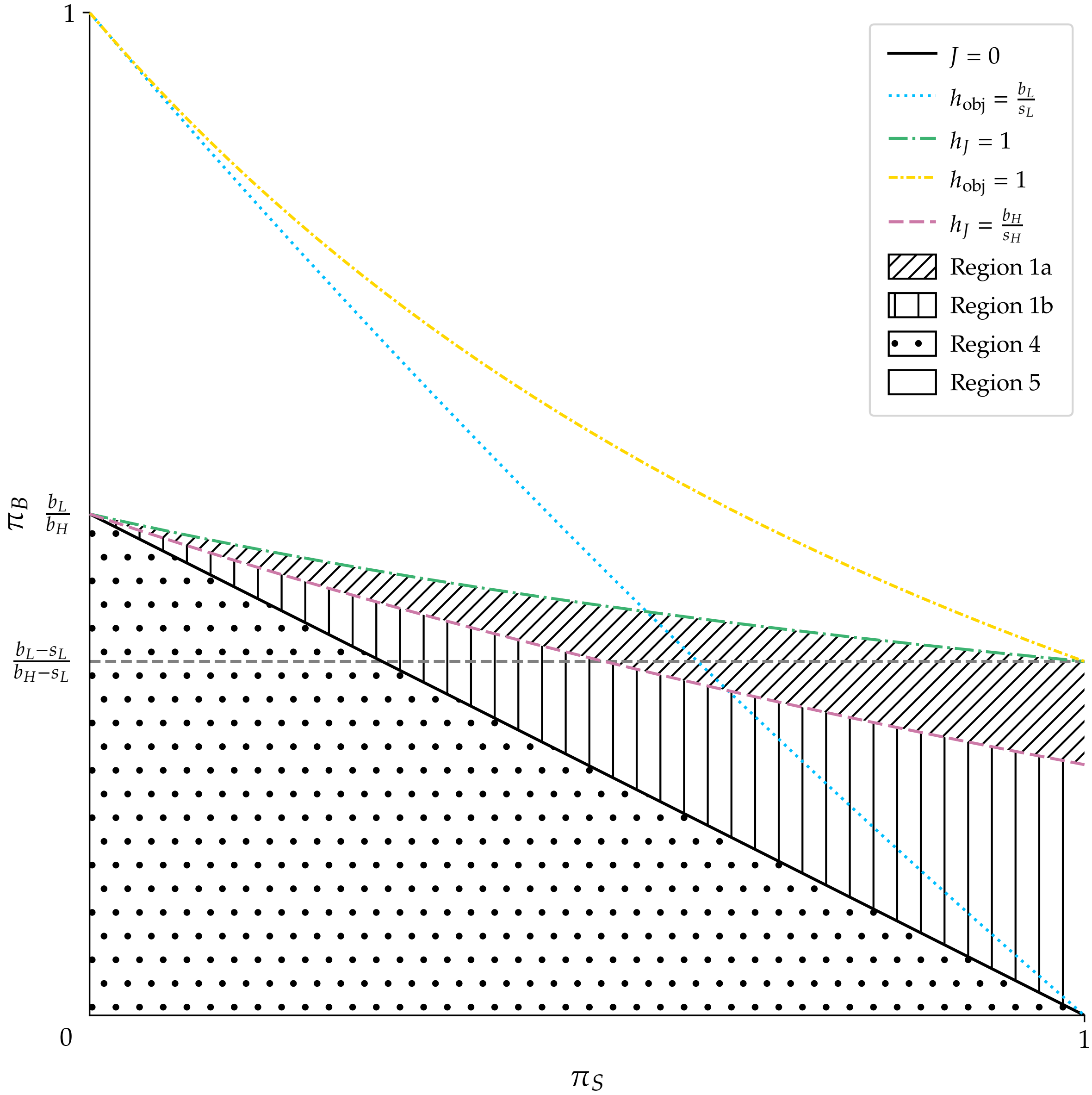

Figure F.2: Four distinct regions in the π_S–π_B space when $\frac{b_H}{s_H} \le \frac{b_L}{s_L}$ and $b_H - s_H > b_L - s_L$. The valuations are set at $b_H = 2.2$, $s_H = 1.5$, $b_L = 1.1$, and $s_L = 0.5$. The black solid, blue dotted, green dash–dot, yellow dash–dot, and magenta dashed lines correspond to $\pi_B(\cdot \mid J = 0)$, $\pi_B(\cdot \mid h_{obj} = b_L/s_L)$, $\pi_B(\cdot \mid h_J = 1)$, $\pi_B(\cdot \mid h_{obj} = 1)$, and $\pi_B(\cdot \mid h_J = b_H/s_H)$, respectively. Note that Region 1b disappears when $\frac{b_H}{s_H} = \frac{b_L}{s_L}$ because $\pi_B(\cdot \mid J = 0)$ and $\pi_B(\cdot \mid h_J = b_H/s_H)$ coincide in this case.

References

Bester, Helmut and Roland Strausz (2000) "Imperfect commitment and the revelation principle: the multi-agent case," *Economics Letters*, Vol. 69, No. 2, pp. 165–171.

——— (2001) "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, Vol. 69, No. 4, pp. 1077–1098.

——— (2007) "Contracting with imperfect commitment and noisy communication," *Journal of Economic Theory*, Vol. 136, No. 1, pp. 236–259.

Čopič, Jernej and Clara Ponsatí (2008) "Robust Bilateral Trade and Mediated Bargaining," *Journal of the European Economic Association*, Vol. 6, No. 2-3, pp. 570–580.

Doval, Laura and Vasiliki Skreta (2022) "Mechanism Design With Limited Commitment," *Econometrica*, Vol. 90, No. 4, pp. 1463–1500.

——— (2024) "Optimal mechanism for the sale of a durable good," *Theoretical Economics*, Vol. 19, No. 2, pp. 865–915.

Eilat, Ran and Ady Pauzner (2021) "Bilateral trade with a benevolent intermediary," *Theoretical Economics*, Vol. 16, No. 4, pp. 1655–1714.

Fanning, Jack (2021) "Mediation in Reputational Bargaining," *American Economic Review*, Vol. 111, No. 8, pp. 2444–72.

——— (2023) "Optimal Dynamic Mediation," *Journal of Political Economy*, Vol. 131, No. 11, pp. 2969–3002.

Forges, Francoise (1986) "An Approach to Communication Equilibria," *Econometrica*, Vol. 54, No. 6, pp. 1375–1385.

Ganguly, Chirantan and Indrajit Ray (2023) "Simple Mediation in a Cheap-Talk Game," *Games*, Vol. 14, No. 3.

Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani (2009) "Mediation, arbitration and negotiation," *Journal of Economic Theory*, Vol. 144, No. 4, pp. 1397–1420.

Gottardi, Piero and Claudio Mezzetti (2024) “Shuttle diplomacy,” *Journal of Economic Theory*, Vol. 216, p. 105794.

Hörner, Johannes, Massimo Morelli, and Francesco Squintani (2015) “Mediation and Peace,” *The Review of Economic Studies*, Vol. 82, No. 4, pp. 1483–1501.

Jarque, Xavier, Clara Ponsatí, and József Sákovics (2003) “Mediation: incomplete information bargaining with filtered communication,” *Journal of Mathematical Economics*, Vol. 39, No. 7, pp. 803–830.

Kamin, Debra (2024) “Buying or Selling a Home? The Rules Have Changed.,” URL: https://www.nytimes.com/2024/08/16/realestate/home-buyers-sellers-rules.html, accessed August 5, 2025.

Kydd, Andrew H. (2003) “Which Side Are You On? Bias, Credibility, and Mediation,” *American Journal of Political Science*, Vol. 47, No. 4, pp. 597–611.

——— (2006) “When Can Mediators Build Trust?” *American Political Science Review*, Vol. 100, No. 3, pp. 449–462.

Loertscher, Simon and Leslie M. Marx (2022) “Incomplete Information Bargaining with Applications to Mergers, Investment, and Vertical Integration,” *American Economic Review*, Vol. 112, No. 2, pp. 616—49.

Lomys, Niccolò and Takuro Yamashita (2022) “A Mediator Approach to Mechanism Design with Limited Commitment,” Working Paper.

Matsuo, Toshihide (1989) “On incentive compatible, individually rational, and ex post efficient mechanisms for bilateral trading,” *Journal of Economic Theory*, Vol. 49, No. 1, pp. 189–194.

Myerson, Roger B. (1981) “Optimal Auction Design,” *Mathematics of Operations Research*, Vol. 6, No. 1, pp. 58–73.

——— (1982) “Optimal coordination mechanisms in generalized principal-agent problems,” *Journal of Mathematical Economics*, Vol. 10, No. 1, pp. 67–81.

——— (1986a) "Acceptable and predominant correlated equilibria," *International Journal of Game Theory*, Vol. 15, No. 3, pp. 133–154.

——— (1986b) "Multistage Games with Communication," *Econometrica*, Vol. 54, No. 2, pp. 323–358.

Myerson, Roger B. and Mark A. Satterthwaite (1983) "Efficient mechanisms for bilateral trading," *Journal of Economic Theory*, Vol. 29, No. 2, pp. 265–281.

Sugaya, Takuo and Alexander Wolitzky (2021) "The Revelation Principle in Multistage Games," *The Review of Economic Studies*, Vol. 88, No. 3, pp. 1503–1540.