THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Yusuke YAMAGUCHI

Le 23 septembre 2025

Essais de théorie économique

École doctorale : **TSE Toulouse Sciences Economiques**
Spécialité : **Sciences Economiques - Toulouse**
Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur David MARTIMORT*

Composition du jury
Rapporteure : Mme Catherine BOBTCHEFF
Rapporteur : M. Joaquin POBLETE
Examinateur : M. Takuro YAMASHITA
Directeur de thèse : M. David MARTIMORT
Président : M. Andrea ATTAR

Essays in Economic Theory

Yusuke Yamaguchi

July 29, 2025

Acknowledgement

I am deeply grateful to my current and former advisors at Toulouse School of Economics, David Martimort and Takuro Yamashita, for their patience, invaluable guidance, and continuous support throughout my studies.

I would also like to thank the many people I have met during my time in Japan and France for their support at various stages of this journey. The list includes, but is not limited to, Milo Bianchi, Fabrice Collard, Daisuke Hirata, Johannes Hörner, Dakang Huang, Taisuke Imai, Akifumi Ishihara, Doh-Shin Jeon, Nour Meddahi, Shintaro Miura, Yasuyuki Miyahara, Nozomu Muto, Akira Okada, Hien Pham, François Salanié, Anna Sanktjohanser, Sara Shahanaghi, Alex Smolin, Norio Takeoka, and Takashi Ui.

Last but not least, I thank my family, Kaori and Haruhisa, whose unwavering support has been vital throughout my life.

Abstract

This thesis comprises three chapters. Chapter 1 analyzes bilateral bargaining mediated by an intermediary who is biased—sharing interests with one party—and lacks both commitment and enforcement power. By comparing the upper bounds of achievable expected social surplus in the mediated bargaining game and in the seller-offer bargaining game, I show that, in the second-best scenario, a biased intermediary can yield a higher expected surplus in equilibrium, provided the cost of employing the intermediary is sufficiently small. This result offers a theoretical rationale for the widespread use of biased intermediaries in practice, even when their bias is common knowledge.

Chapter 2 examines a standard moral hazard setting in which the agent exhibits dual risk-aversion—that is, he overweights the probability of bad outcomes relative to objective probabilities. I show that, under some regularity conditions, the cost-minimizing contract for each inducible action takes the form of a debt contract: the wage is constant up to a certain output level and then increases at the same rate as output thereafter. Accordingly, the result offers a new explanation for the prevalence of simple linear contracts in practice.

Chapter 3, which is based on the joint work with Takuro Yamashita, studies a stylized model of adverse selection. I show that under certain conditions, the unique equilibrium outcome is no trade, regardless of the richness of the soft information structure. In contrast, when hard information is available, there exists an equilibrium in which trade occurs and the seller obtains a positive expected payoff. Taken together, these results underscore the importance of explicitly accounting for the availability of hard information when considering informationally robust predictions.

Contents

Chapter 1

Bilateral Bargaining with a Biased Intermediary

1.1 Introduction

Bilateral bargaining is often facilitated by an *intermediary* who communicates with both negotiating parties and offers them a resolution. Such mediation obfuscates the private information that would otherwise be revealed through direct communication, and is thus believed to mitigate strategic incentives in communication and help achieve 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, which can harm 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. Because she earns a commission proportional to the sale price, she has

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

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 chapter 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 wants to trade it with the buyer. Each trader's valuation of the good is his private information and is binary and independently and distributed. There is always a gain 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 commission proportional to the price, and the intermediary's payoff is the sum of these commissions. If no trade occurs, all players get zero.

To characterize the bounds of achievable outcomes when the players are allowed to make any 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).[3] 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

[2] In the U.S. Realtors industry, it has been standard for the seller to pay commissions to both his agent and the buyer's agent. Traditionally, the seller pays about 5% of the sale price as commission, which is then split equally between the two agents. Recently, this rule has 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).

[3] In the seller-offer bargaining game, a mediation plan recommends a price to the seller and a maximum acceptable price to the buyer.

that price; if at least one of them rejects it, no trade occurs.[4]

The set of incentive-compatible mediation plans, known as *communication equilibrium* (*CE*), characterizes the set of achievable outcomes under *some* communication (see Forges, 1986). However, some CEs are not robust under standard trembling-hand arguments and are therefore considered unrealistic. 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 chapter is that if the intermediary's bias is sufficiently small, the mediated bargaining game can yield a higher ex-ante expected social surplus than the seller-offer bargaining game in the second-best case. I establish this by deriving necessary conditions for acceptable CEs in both games. Roughly, these conditions imply that the player who makes an offer always proposes either of the buyer's maximum acceptable prices. Intuitively, since traders accept any mutually acceptable price under acceptable CE, the player with bargaining power has the incentive to offer as high a price as possible. The necessary conditions imply that whenever there is a gain trade, there are only two possible prices (low or high price) that can be offered in equilibrium. Thus, the players' IC constraints can be written as linear inequalities in the total probabilities that the high price is offered. 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-type buyer's maximum acceptable price whenever there is a gain trade. In the mediated bargaining game, the solution implies a threshold for the traders' payment/revenue ratio, above which the SB outcome shifts regime. This ratio, which is increasing in the commission rates, represents the degree of the intermediary's bias. If the ratio is below the threshold, the intermediary offers the low price to the pair of high-type traders, knowing that it can be rejected. Intuitively, this is possible because,

[4] In the seller-offer bargaining game, only the buyer responds to the seller's offer, and trade occurs if and only if he accepts.

not knowing the seller's private information, the intermediary can be incentivized to offer the low price. As the ratio increases, the probability of trade and hence the expected social surplus decreases. Above the threshold, the intermediary is no longer incentivized to offer the low price and always offers the high-type buyer's maximum acceptable price, replicating the SB outcome in the seller-offer bargaining game. Since the expected surplus is higher in the first regime, this shows that when the commission cost is sufficiently small, her presence improves efficiency relative to having no intermediary at all. This result justifies the existence of biased intermediaries in practice, even when their bias is common knowledge.

Finally, it is worth mentioning that the mediated bargaining game studied here represents a minimal departure from the seller-offer game in the sense that the intermediary is "weak." She has exactly the same instrument as the seller in the seller-offer bargaining game; that is, she merely offers a price and cannot enforce binding contracts or mechanisms. She also has no private information and hence is considered to have no expertise. Yet, the main result demonstrates that even such a weak intermediary can be beneficial.

1.1.1 Related literature

Three features of the intermediary—bias, lack of commitment, and lack of enforcement—distinguish the present work from existing papers. 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 it is natural to study a "weak" intermediary as a minimal departure from bilateral bargaining in examining the effect of biased mediation, much of the literature takes Myerson and Satterthwaite (1983)

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

[6] In this sense, "no enforcement" is equivalent to imposing ex post IR constraints rather than ex ante ones.

as the benchmark 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, 2021). Relatedly, Eilat and Pauzner (2021) study bilateral trade with a benevolent intermediary who lacks commitment power.[7] The novelty of the present work is thus to consider all three features simultaneously—bias, lack of commitment, and lack of enforcement—which provides a natural way to study a minimal departure from bilateral bargaining.[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. This result resonates with the main result of this chapter that a biased intermediary can help achieve a more efficient outcome.[9]

The present work 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 private information inherent in offers and counteroffers.[10] See Čopič and Ponsatí (2008); Fanning (2021, 2023); Jarque, Ponsatí and Sákovics (2003)

[7]They directly analyze the game between the traders and the intermediary, in which the intermediary offers a mechanism. Note that the intermediary in my model merely offers a price and is not allowed to design a mechanism. In this regard, 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 (2021) and Doval and Skreta (2024) analyze seller-optimal mechanisms under limited commitment.

[9]It is also worth noting that a biased intermediary does not help achieve ex post efficiency in the present work (see Proposition 1.1 and Footnote 24).

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

for studies focusing on this facilitative role, and Ganguly and Ray (2012); Goltsman, Hörner, Pavlov and Squintani (2009); Hörner, Morelli and Squintani (2015) for other work on mediation.

Methodologically, the present work 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 chapter, the communication revelation principle implies that the set of CEs characterizes the set of outcomes that can be achieved under some preplay and intraplay communication between the players. In particular, it is without loss of generality to focus on the canonical communication devices described above, which take the players' types as inputs and output action recommendations.

The remainder of the chapter is organized as follows. Section 1.2 describes the model and defines CE and acceptable CE. Section 1.3 presents some preliminary results. Section 1.4 provides the main result. Section 1.5 discusses alternative payoff specifications and concludes the chapter. Some proofs are provided in Appendix.

1.2 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 trade for all type profiles except (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 1.2.1), 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 1.2.2), 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. I assume $h \leq \min\left\{\frac{b_L}{s_L}, \frac{b_H}{s_H}\right\}$, so that a *mutually acceptable* price exists if and only if there is a gain trade.[11] Since the intermediary's payoff is increasing in the price p, she shares interests with the seller and is thus considered *biased*.

1.2.1 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 *mediation plan*, which is a probability distribution $\mu^{\text{SO}} \in \Delta(Q \times R^{\text{SO}})$ and is common knowledge among the players.

The game proceeds as follows:

[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; that is, $p \in \left[\frac{s}{1-\delta_S}, \frac{b}{1+\delta_B}\right]$. Provided $h \leq \min\left\{\frac{b_L}{s_L}, \frac{b_H}{s_H}\right\}$, this interval is nonempty for all type profiles except (s_H, b_L). See also Footnote 17.

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

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 $\left(q, r^{\mathrm{SO}}\right) \in Q \times R^{\mathrm{SO}}$ is drawn with probability $\mu^{\mathrm{SO}}\left(q, r^{\mathrm{SO}}\right)$, but the players do not observe which one is drawn.
4. The communication device privately recommends a price $q\left(\tilde{s}, \tilde{b}\right)$ to the seller and a maximum acceptable price $r^{\mathrm{SO}}\left(\tilde{s}, \tilde{b}\right)$ 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.

Remark 1.1. I can consider a more general game in which the communication device either *sequentially* recommends actions—recommending a price at Time 4 and a response after Time 5—or recommends a response *for each possible price* at Time 4, instead of recommending only a maximum acceptable price. However, these generalizations do not affect the results, as long as the analysis is restricted to "acceptable communication equilibria," which will be defined shortly. The same remark applies to the mediated bargaining game.

1.2.2 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 now recommends a price to the intermediary, along with a maximum acceptable price for the buyer and a minimum acceptable price for the seller. Formally, the response recommendation now maps into $\mathbb{R}_+^2$; that is, $r^{\mathrm{MB}}(s, b) = \left(r_S^{\mathrm{MB}}(s, b), r_B^{\mathrm{MB}}(s, b)\right)$, where r_S^{MB} and r_B^{MB} 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 $(q, r^{\mathrm{MB}}) \in Q \times R^{\mathrm{MB}}$ is drawn with probability $\mu^{\mathrm{MB}}(q, r^{\mathrm{MB}})$, 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 $r_S^{\mathrm{MB}}(\tilde{s}, \tilde{b})$ to the seller, and a maximum acceptable price $r_B^{\mathrm{MB}}(\tilde{s}, \tilde{b})$ 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.

1.2.3 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. 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 $r^{\mathrm{SO}}(s, b) = p$. 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 $\tilde{p} \in (p, b)$, the buyer is better off accepting it, implying that obedience to (q, r^{SO}) is not robust to such trembles.[13] By contrast, an acceptable CE is robust to such perturbations.[14]

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:

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

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 $\sigma_{S1}^{\text{SO}} : \Theta_S \to \Theta_S$ is a manipulation in report and $\sigma_{S2}^{\text{SO}} : \Theta_S \times \mathbb{R}_+ \to \mathbb{R}_+$ 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

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

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

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

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 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 $r^{\text{SO}*}\left(s, \tilde{b}(b)\right) = \tilde{b}(b)$, and accepts an offer if and only if the price is smaller than or equal to $p(b) \equiv \sigma_{B2}^{\text{SO}}\left(b, \tilde{b}(b)\right)$.[15] The buyer's ex ante expected payoff from such manipulation is thus

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

A mediation plan $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$ is an acceptable CE if no player has a profitable manipulation.[16]

Definition 1.1. 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 $\sigma_i^{\text{SO}} \in \Sigma_i^{\text{SO}}$,

$$\sum_{q\in Q} \mu^{\text{SO}}(q) U_i^{\text{SO}}(q) \ge \sum_{q\in Q} \mu^{\text{SO}}(q) U_i^{\text{SO}}(q \circ \sigma_i^{\text{SO}}). \tag{1.1}$$

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 $p_B(b) \equiv \frac{b}{1+\delta_B}$.[17] Hence, the response recommendation described at the beginning of this section is captured by $r^{\text{MB}*}$, defined as $r^{\text{MB}*}(s,b) = \left(p_S(s), p_B(b)\right)$

[15] 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.

[16] Note that in this game, every acceptable CE is a *sequential communication equilibrium* (SCE). Myerson (1986b) defines 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 1.1 further implies that no acceptable CE recommends a codominated action to the seller either, thereby establishing the equivalence between CE and SCE (see Footnote 18). A similar argument can be made for the mediated bargaining game.

[17] Hence, any price $p \in \left[p_S(s), p_B(b)\right]$ is mutually acceptable for a type-s seller and a type-b buyer. Since $h \in \left(1, \min\left\{\frac{b_L}{s_L}, \frac{b_H}{s_H}\right\}\right]$ 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 trade.

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 $\left(q, r^{\mathrm{MB}*}\right)$ 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

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

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 $\left(q, r^{\mathrm{MB}*}\right)$ using $\sigma_I \in \Sigma_I$ while the traders are honest and obedient, her ex ante expected payoff is

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

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

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

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

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

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

Definition 1.2. 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:

1. For all $\sigma_I \in \Sigma_I$,
$$\sum_{q\in Q} \mu^{\text{MB}}(q)V(q) \ge \sum_{q\in Q} \mu(q)V(\sigma_I \circ q); \tag{1.2}$$

2. For all $i \in \{S, B\}$ and all $\sigma_i^{\text{MB}} \in \Sigma_i^{\text{MB}}$,
$$\sum_{q\in Q} \mu^{\text{MB}}(q)U_i^{\text{MB}}(q) \ge \sum_{q\in Q} \mu(q)U_i^{\text{MB}}(q \circ \sigma_i^{\text{MB}}). \tag{1.3}$$

Remark 1.2. Strictly speaking, the communication revelation principle is established only for finite games. To ensure the validity of the analysis, it is therefore necessary to have finite sets of possible prices in both games. Discretizing the set of prices yields qualitatively the same results. For clarity and brevity, however, I adopt a continuous formulation throughout the chapter.

Remark 1.3. In both games, the players' ex ante incentive compatibility (IC) constraint—namely, (1.1) in the seller-offer bargaining game and (1.2) and (1.3) 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, I can construct an ex ante manipulation that prescribes deviation only in that scenario and otherwise follows the recommendation. 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 establishes the equivalence between ex ante and interim IC constraints.

1.3 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 $r^{\text{SO}*}$ or $r^{\text{MB}*}$—this provides an incentive to offer the highest acceptable price to the buyer whenever there is a gain 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.

1.3.1 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.[18]

Lemma 1.1. *A mediation plan $\mu^{\text{SO}} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$ 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 $q \in \text{supp}\big(\mu^{\text{SO}}\big)$,*

$$\begin{aligned} &q(s_H, b_H) = b_H, \\ &q(s_L, b) \in \{b_L, b_H\} \text{ for all } b \in \Theta_B. \end{aligned} \tag{1.4}$$

Proof. I prove the contraposition. Consider a mediation plan $\mu^{\text{SO}} \in \Delta\big(Q \times \{r^{\text{SO}*}\}\big)$ that violates (1.4) for some $q \in \text{supp}\big(\mu^{\text{SO}}\big)$. I show that the seller has a profitable

[18] 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.

manipulation. If type s seller reports $\tilde{s}$ and receives a recommendation p, his posterior belief that the buyer is high type is

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

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

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

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 $\nu^{\mathrm{SO}}(b_H \mid s_H, p) > 0$. 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 $\nu^{\mathrm{SO}}(b_H \mid s_L, p) \geq 0$. If $p \in (b_L, b_H)$ and $\nu^{\mathrm{SO}}(b_H \mid s_L, p) > 0$, 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 $\mu^{\mathrm{SO}} \in \Delta\big(Q \times \{r^{\mathrm{SO}*}\}\big)$, construct a modified mediation plan $\mu^{\mathrm{SO}*}$ as follows. For each $q \in \mathrm{supp}\big(\mu^{\mathrm{SO}}\big)$, define $q^{\mathrm{SO}*} \in Q$ by

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

Define $\mu^{\mathrm{SO}*} \in \Delta\big(Q \times \{r^{\mathrm{SO}*}\}\big)$ by setting $\mu^{\mathrm{SO}*}\big(q^{\mathrm{SO}*}\big) = \mu^{\mathrm{SO}}(q)$. That is, $\mu^{\mathrm{SO}*}$ 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 $\mu^{\mathrm{SO}*}$ make it without loss of generality to focus on the set of such mediation plans. First, each $q^{\mathrm{SO}*}$ 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.[19] This leads to the following lemma.

Lemma 1.2. *For any mediation plan $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$, 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 1.1, consider an arbitrary mediation plan $\mu^{\text{SO}} \in \Delta(Q \times \{r^{\text{SO}*}\})$ that satisfies (1.4) 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 $x^{\text{SO}}_{\text{HL}}$ and $x^{\text{SO}}_{\text{LL}}$ 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:

$$x^{\text{SO}}_{\text{HL}} = \sum_{q:\, q(s_H, b_L) = b_H} \mu^{\text{SO}}(q),$$

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

Note that the latter probability is identical 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

[19] 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.

payoff from misreporting is thus at most

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

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

$$\max\{\pi_B(b_H - s_L), b_L - s_L\} \tag{1.5}$$

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 $\mu^{\text{SO}*}$.

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 $(1-\pi_S)(1-x^{\text{SO}}_{\text{LL}})$. 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 $\mu^{\text{SO}*}$, 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

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

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

$$(1-\pi_S)(1-x^{\text{SO}}_{\text{LL}})(b_H - b_L). \tag{1.6}$$

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

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

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

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

IC constraints

By Lemma 1.2, it suffices to focus on $\mu^{\text{SO}*}$, which is constructed from some mediation plan $\mu^{\text{SO}} \in \Delta(Q\times\{r^{\text{SO}*}\})$ that satisfies (1.4) for all $q \in \text{supp}(\mu^{\text{SO}})$. As in the proof of Lemma 1.2, define $x^{\text{SO}}_{\text{LH}}$ as the total probability under μ^{SO} that the price b_H is recommended for the report (s_L, b_H):

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

As in the case of $x^{\text{SO}}_{\text{LL}}$, this probability is identical 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

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

Combined with (1.5), he has no incentive to misreport his type if

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

Now, suppose that he reports his type truthfully and receives the recommendation b_H. His posterior belief that the buyer is high type is $\frac{\pi_B x_{\mathrm{LH}}^{\mathrm{SO}}}{\pi_B x_{\mathrm{LH}}^{\mathrm{SO}} + (1-\pi_B)x_{\mathrm{LL}}^{\mathrm{SO}}}$. Hence, he follows the recommendation if [20]

$$\begin{aligned}&\frac{\pi_B x_{\mathrm{LH}}^{\mathrm{SO}}}{\pi_B x_{\mathrm{LH}}^{\mathrm{SO}} + (1 - \pi_B)x_{\mathrm{LL}}^{\mathrm{SO}}}(b_H - s_L) \geq b_L - s_L\\ \iff &\pi_B x_{\mathrm{LH}}^{\mathrm{SO}}(b_H - b_L) \geq (1 - \pi_B)x_{\mathrm{LL}}^{\mathrm{SO}}(b_L - s_L).\end{aligned} \tag{SO-IC$_{s_L}$-1}$$

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

$$\begin{aligned}&b_L - s_L \geq \frac{\pi_B(1 - x_{\mathrm{LH}}^{\mathrm{SO}})}{\pi_B(1 - x_{\mathrm{LH}}^{\mathrm{SO}}) + (1 - \pi_B)(1 - x_{\mathrm{LL}}^{\mathrm{SO}})}(b_H - s_L)\\ \iff &(1 - \pi_B)(1 - x_{\mathrm{LL}}^{\mathrm{SO}})(b_L - s_L) \geq \pi_B(1 - x_{\mathrm{LH}}^{\mathrm{SO}})(b_H - b_L).\end{aligned} \tag{SO-IC$_{s_L}$-2}$$

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 (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

[20] As discussed in the proof of Lemma 1.2, it suffices to show that he prefers offering b_H over b_L. Similar argument applies to the recommendation b_L.

payoff of $(1-\ \pi_S)(1-\ x^{\text{SO}}_{\text{LH}})(b_H - b_L)$. Combined with (1.6), his IC constraint is

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

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.

1.3.2 Mediated bargaining game

For the mediated bargaining game, analogs of Lemmas 1.1 and 1.2 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 trade.[21] The key difference from the seller-offer bargaining game is that the recommendation for the report (s_H, b_H) is not necessarily the higher 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 1.3. *A mediation plan $\mu^{\text{MB}} \in \Delta(Q \times \{r^{\text{MB}*}\})$ is an acceptable CE only if it always recommends either $p_B(b_L)$ or $p_B(b_H)$ whenever there is a gain trade. That is, for all $q \in \text{supp}(\mu^{\text{MB}})$,*

$$q(s,b) \in \{p_B(b_L), p_B(b_H)\} \quad \textit{for all } (s,b) \in \Theta_S \times \Theta_B \setminus \{(s_H, b_L)\}. \tag{1.7}$$

Proof. See Appendix A.1. □

Intuitively, when the traders follow $r^{\text{MB}*}$, trade occurs if and only if the intermediary offers a mutually acceptable price. Thus, she clearly has an incentive to offer the highest acceptable price—either $p_B(b_L)$ or $p_B(b_H)$. Therefore, if a mediation plan were

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

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 $p_{\text{HL}} \in (p_B(b_H), +\infty)$, and for any mediation plan $\mu^{\text{MB}} \in \Delta(Q \times \{r^{\text{MB}*}\})$, construct a modified mediation plan $\mu^{\text{MB}*}$ as follows. For each $q \in \text{supp}(\mu^{\text{MB}})$, define $q^{\text{MB}*} \in Q$ by

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

Define $\mu^{\text{MB}*}$ by setting $\mu^{\text{MB}*}(q^{\text{MB}*}) = \mu^{\text{MB}}(q)$. That is, $\mu^{\text{MB}*}$ is obtained by a similar modification as in the case of $\mu^{\text{SO}*}$, except that the recommendation for (s_H, b_L) is now replaced with p_{HL}, which is not mutually acceptable for any trader pair. Compared to the original mediation plan, the modified plan $\mu^{\text{MB}*}$ 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 1.4. *For any mediation plan $\mu^{\text{MB}} \in \Delta(Q \times \{r^{\text{MB}*}\})$, if μ^{MB} is an acceptable CE, then so is $\mu^{\text{MB}*}$. Moreover, μ^{MB} and $\mu^{\text{MB}*}$ yield the same ex ante expected social surplus when all players are honest and obedient.*

Proof. See Appendix A.2. □

IC constraints

By Lemma 1.4, it suffices to focus on $\mu^{\text{MB}*}$, which is constructed from some mediation plan $\mu^{\text{MB}} \in \Delta(Q \times \{r^{\text{MB}*}\})$ that satisfies (1.7) for all $q \in \text{supp}(\mu^{\text{MB}})$. As in the seller offer bargaining game, let $x^{\text{MB}}_{\text{HH}}$, $x^{\text{MB}}_{\text{LH}}$, and $x^{\text{MB}}_{\text{LL}}$ 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:

$$x^{\text{MB}}_{\text{HH}} = \sum_{q:\, q(s_H, b_H) = p_B(b_H)} \mu^{\text{MB}}(q),$$
$$x^{\text{MB}}_{\text{LH}} = \sum_{q:\, q(s_L, b_H) = p_B(b_H)} \mu^{\text{MB}}(q),$$
$$x^{\text{MB}}_{\text{LL}} = \sum_{q:\, q(s_L, b_L) = p_B(b_H)} \mu^{\text{MB}}(q).$$

These probabilities are identical under μ^{MB} and $\mu^{\text{MB}*}$ because the relevant recommendations remain unchanged. Then, the players' IC constraints under $\mu^{\text{MB}*}$ can be expressed as linear inequalities in $x^{\text{MB}}_{\text{HH}}$, $x^{\text{MB}}_{\text{LH}}$, and $x^{\text{MB}}_{\text{LL}}$, 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 [22]

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

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

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

[22]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.

Thus, he has no incentive to misreport if

$$\pi_B\left[x^{\text{MB}}_{\text{LH}}\frac{b_H}{h}+\left(1-x^{\text{MB}}_{\text{LH}}\right)\frac{b_L}{h}-s_L\right]+(1-\pi_B)(1-x^{\text{MB}}_{\text{LL}})\left(\frac{b_L}{h}-s_L\right)$$
$$\geq \pi_B\left[x^{\text{MB}}_{\text{HH}}\frac{b_H}{h}+\left(1-x^{\text{MB}}_{\text{HH}}\right)\frac{b_L}{h}-s_L\right]$$
$$\iff (1-\pi_B)(1-x^{\text{MB}}_{\text{LL}})(b_L-hs_L)\geq \pi_B(x^{\text{MB}}_{\text{HH}}-x^{\text{MB}}_{\text{LH}})(b_H-b_L). \tag{MB-IC$_{s_L}$}$$

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

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

Thus, he has no incentive to misreport if

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

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:

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

where $D_{b_H} = \pi_S \pi_B x^{\text{MB}}_{\text{HH}} + (1-\pi_S)\pi_B x^{\text{MB}}_{\text{LH}} + (1-\pi_S)(1-\pi_B) x^{\text{MB}}_{\text{LL}}$ 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 $p_B(b_L)$:

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

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

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

where $D_{b_L} = \pi_S \pi_B (1 - x^{\text{MB}}_{\text{HH}}) + (1-\pi_S)\pi_B (1 - x^{\text{MB}}_{\text{LH}}) + (1-\pi_S)(1-\pi_B)(1 - x^{\text{MB}}_{\text{LL}})$ 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 $p_B(b_H)$:

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

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

1.4 Bound on the Ex Ante Expected Social Surplus

As the IC constraints derived in Sections 1.3.1 and 1.3.2, as well as the ex ante expected social surplus, are linear in $x^{\text{SO}} = (x^{\text{SO}}_{\text{LH}}, x^{\text{SO}}_{\text{LL}})$ or $x^{\text{MB}} = (x^{\text{MB}}_{\text{HH}}, x^{\text{MB}}_{\text{LH}}, x^{\text{MB}}_{\text{LL}})$, 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 chapter. 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.[23]

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; otherwise, he would deviate and offer the high price 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. But 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, which depends on the ratio $h = \frac{1+\delta_B}{1-\delta_S}$ in the mediated bargaining game. A formal analysis (omitted for brevity) leads to the following proposition.[24]

Proposition 1.1. *There exists an acceptable CE that achieves ex post efficiency under the following 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* $\pi_B \leq \frac{b_L - hs_L}{b_H - hs_L}$.

[23]That is, the candidate mediation plans are $x^{\text{SO}} = (0, 0)$ in the seller-offer bargaining game and $x^{\text{MB}} = (1, 0, 0)$ 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.

[24]I can also consider a purely benevolent intermediary who takes no commissions and aims to maximize, say, the expected trade gain. 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 individual rationality is satisfied even in Matsuo (1989), so no enforcement power is required.

Note that the condition $\pi_B \leq \frac{b_L - h s_L}{b_H - h s_L}$ can also be interpreted as imposing an upper bound on the ratio h, denoted by $\bar{h}^{\text{Eff}}$:[25]

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

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 1.1.

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 1.2, for any $\mu^{\text{SO}*}$, the ex ante expected social surplus is given by

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

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

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

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

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,

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

the seller always offers the price b_H whenever there is a gain from trade. This leads to the following proposition.

Proposition 1.2. *If $\pi_B > \frac{b_L - s_L}{b_H - s_L}$, in the unique acceptable CE of the seller-offer bargaining game, the traders (s_H, b_H) and (s_L, b_H) trade at the price b_H with probability one, while (s_L, b_L) do not trade with probability one.*

Proof. See Section A.3. □

As discussed in Section 1.3.1, 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 this price 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.

1.4.2 Second-best outcome in the mediated bargaining game

Next, consider the mediated bargaining game. Assume $h > \bar{h}^{\mathrm{Eff}}$ so that ex post efficiency cannot be achieved in any acceptable CE. For any $\mu^{\mathrm{MB}*}$, the ex ante expected social surplus is given by [26]

$$\begin{aligned} &\sum_{(s,b)\in\Theta_S\times\Theta_B} \Pr(s,b) \sum_{q\in Q} \mu^{\mathrm{MB}*}(q)(b-s)\cdot \mathbf{1}_{\{p_S(s)\le q(s,b)\le p_B(b)\}} \\ &= \pi_S\pi_B(b_H - s_H)x^{\mathrm{MB}}_{\mathrm{HH}} + (1-\pi_S)\pi_B(b_H - s_L) \\ &\quad + (1-\pi_S)(1-\pi_B)(b_L - s_L)(1 - x^{\mathrm{MB}}_{\mathrm{LL}}). \end{aligned} \tag{1.8}$$

Thus, the relevant linear program is

$$\begin{aligned} &\max_{x^{\mathrm{MB}}\in[0,1]^3} \ (1.8) \\ &\text{subject to} \quad (\text{MB-IC}_{s_H}), (\text{MB-IC}_{s_L}), (\text{MB-IC}_{b_H}), (\text{IC}_{\text{Int}}\text{-}1), \text{ and } (\text{IC}_{\text{Int}}\text{-}2). \end{aligned} \tag{MB-P}$$

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

Note that, by construction, trade occurs with probability one for the pair (s_L, b_H). Thus, the trade-off lies between the expected social surplus $\pi_S\pi_B(b_H - s_H)x^{\mathrm{MB}}_{\mathrm{HH}}$ generated by the high-type pair (s_H, b_H) and $(1-\pi_S)(1-\pi_B)(b_L - s_L)(1 - x^{\mathrm{MB}}_{\mathrm{LL}})$ generated by the low-type pair (s_L, b_L).

For the sake of exposition, I focus on the case where the primitives satisfy the following four conditions:

C1. $b_H - s_H > b_L - s_L$;

C2. $\frac{b_H}{s_H} > \frac{b_L}{s_L}$;

C3. $\pi_B \geq \frac{(1-\pi_S)b_L}{b_H}$; and

C4. $\pi_B < \min\left\{\pi_B^{h_J}(\pi_S), \pi_B^{h_{J-K}}(\pi_S)\right\}$,

where

$$\pi_B^{h_J}(\pi_S) = \frac{\pi_S b_H(b_L - s_L) + (1-\pi_S)(b_H - b_L)b_L}{b_H[\pi_S(b_H - s_L) + (1-\pi_S)(b_H - b_L)]},$$
$$\pi_B^{h_{J-K}}(\pi_S) = \frac{(b_L - s_L)[\pi_S(b_H - s_H) + (1-\pi_S)(b_H - b_L)]}{\pi_S(b_H - s_H)(b_H - s_L) + (1-\pi_S)(b_H - b_L)(b_L - s_L)}.$$

C1 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). C2 is a technical condition that implies $\min\left\{\frac{b_L}{s_L}, \frac{b_H}{s_H}\right\} = \frac{b_L}{s_L}$, which is the upper bound on the ratio h below which a mutually acceptable price exists if and only if there is a gain from trade.[27] C3 implies that the intermediary has a stronger incentive to deviate when the recommended price is $p_B(b_L)$ rather than $p_B(b_H)$.[28] C4 is also a technical condition that ensures the threshold h^* in the next proposition exceeds one, so that the condition $h \leq h^*$ has substantive meaning. With the exception of one case noted in Footnote 29, the qualitative result remains the same even when any of C1–C3 fails to hold.[29]

[27] See Footnote 17.

[28] See the proof of Proposition 1.3.

[29] The analysis of the other cases is provided in Section A.5. There is one case in which trade occurs with probability less than one for both pairs (s_H, b_H) and (s_L, b_L) if h exceeds the threshold. See Case 5 therein for details.

Proposition 1.3. *Assume $h > \bar{h}^{\text{Eff}}$ and C1–C4. Then, there exists a threshold h^* for the ratio h such that in the SB outcome:*

1. *If $h \leq h^*$, the traders (s_H, b_H) trade at the price $p_B(b_H)$ with probability less than one, while the traders (s_L, b_H) and (s_L, b_L) trade at the price $p_B(b_L)$ with probability one. The associated ex ante expected social surplus is decreasing in h.*

2. *If $h > h^*$, the traders (s_H, b_H) and (s_L, b_H) trade at the price $p_B(b_H)$ with probability one, while the traders (s_L, b_L) do not trade with probability one. The associated ex ante expected social surplus is constant in h.*

Moreover, the ex ante expected social surplus in Case 1 is weakly higher than in Case 2.

Proof. See Section A.4. □

Recall that the unique candidate for an acceptable CE that achieves ex post efficiency is $x^{\text{MB}} = (1, 0, 0)$; that is, always recommending the price $p_B(b_H)$ to the report (s_H, b_H) and the price $p_B(b_L)$ to the reports (s_L, b_H) and (s_L, b_L). This mediation plan is not an acceptable CE when $h > \bar{h}^{\text{Eff}}$ because the low-type seller would misreport to obtain the high price $p_B(b_H)$. Given the expression (1.8), the expected social surplus close to the ex post efficient level may be achieved by reducing $x^{\text{MB}}_{\text{HH}}$, the probability of recommending the high price for (s_H, b_H), to deter the low-type seller from misreporting. Indeed, the first regime in Proposition 1.3 is obtained by keeping $x^{\text{MB}}_{\text{LH}} = x^{\text{MB}}_{\text{LL}} = 0$ and decreasing $x^{\text{MB}}_{\text{HH}}$ from 1 until (MB-IC$_{s_L}$) binds.

Let $x^{\text{MB}*}_{\text{HH}}$ denote the value of $x^{\text{MB}}_{\text{HH}}$ at which (MB-IC$_{s_L}$) binds when $x^{\text{MB}}_{\text{LH}} = x^{\text{MB}}_{\text{LL}} = 0$. Under the mediation plan $\left(x^{\text{MB}*}_{\text{HH}}, 0, 0\right)$, the traders have no incentive to deviate. However, the intermediary may deviate and offer the price $p_B(b_H)$ when the recommendation is $p_B(b_L)$; that is, (IC$_{\text{Int}}$-2) may not be satisfied. Note that she follows the recommendation $p_B(b_L)$ only if she sufficiently believes that the seller is unlikely to be high type, who would reject $p_B(b_L)$. This belief is reinforced as the probability of recommending $p_B(b_L)$ for (s_H, b_H), given by $1 - x^{\text{MB}*}_{\text{HH}}$, becomes smaller, implying a lower bound on $x^{\text{MB}*}_{\text{HH}}$ above which $\left(x^{\text{MB}*}_{\text{HH}}, 0, 0\right)$ is incentive compatible for the intermediary.[30] As can be seen from

[30] Recall that C3 ensures that she follows the recommendation $p_B(b_H)$ whenever she follows the recommendation $p_B(b_L)$.

(MB-IC$_{s_L}$), $x^{\text{MB}*}_{\text{HH}}$ is decreasing in h.[31] Therefore, there exists an upper bound on h below which $(x^{\text{MB}*}_{\text{HH}}, 0, 0)$ is an acceptable CE. Since the equilibrium probability of trade for (s_H, b_H), given by $x^{\text{MB}*}_{\text{HH}}$, is decreasing in h, so does the expected social surplus.

If h exceeds the threshold, $(x^{\text{MB}*}_{\text{HH}}, 0, 0)$ is no longer an acceptable CE because the intermediary would prefer to deviate and offer $p_B(b_H)$. In this case, she believes that the seller is sufficiently likely to be high type, and the gain from higher commissions becomes too large to ignore.[32] As a result, she always offers $p_B(b_H)$ whenever there is a gain from trade. Since the equilibrium trade probability is independent of h, the expected social surplus is constant in h. The expected social surplus is weakly higher in the first regime, where h is below the threshold, indicating that an intermediary with high commission costs may reduce efficiency.

Finally, I compare the SB outcomes in the seller-offer bargaining game and the mediated bargaining game. Propositions 1.2 and 1.3 implies that the SB level of expected social surplus coincides in the two games when the ratio h exceeds the threshold. This is because, when h is large, the intermediary is incentivized to always offer the high price, just as the seller does in the SB outcome of the seller-offer bargaining game. This leads to the following corollary.

Corollary 1.1. *If $\pi_B > \frac{b_L - s_L}{b_H - s_L}$ and $h \leq h^*$, the mediated bargaining game can achieve a higher ex ante expected social surplus in an acceptable CE than the seller-offer bargaining game.*

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

[31] As discussed in the paragraph following Proposition 1.1, an increase in h makes deviation more attractive for the low-type seller because C4 implies $\pi_B < \frac{b_L}{b_H}$. To deter him from misreporting, it is therefore necessary to decrease $x^{\text{MB}}_{\text{HH}}$ so that (MB-IC$_{s_L}$) remains binding.

[32] If C4 fails to hold, then the buyer is sufficiently likely to be high type that the intermediary offers $p_B(b_H)$ even when h is close to one.

1.5 Conclusion

This chapter 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 shares interests with the seller, possesses the same instruments as the seller, and has no private information. 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 $b - p$, 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 chapter, the commission rates were assumed to be exogenously fixed. However, in light of recent reforms in the U.S. real estate industry (see Footnote 2), allowing the traders and the intermediary to bargain over the commission rates would be a natural extension and could contribute to the ongoing policy debate. I plan to explore these extensions in future research.

Appendix to Chapter 1

This appendix is divided into several sections. Section A.1 provides the proof of Lemma 1.3. Section A.2 provides the proof of Lemma 1.4. Section A.3 provides the proof of Proposition 1.2. Section A.4 provides the proof of Proposition 1.3. Section A.5 characterizes the SB outcome in the mediated bargaining game when any of C1–C3 fails to hold.

A.1 Proof of Lemma 1.3

I prove the contraposition. Consider a mediation plan $\mu^{\mathrm{MB}} \in \Delta\big(Q \times \{r^{\mathrm{MB}*}\}\big)$ that violates (1.7) for some $q \in \mathrm{supp}\big(\mu^{\mathrm{MB}}\big)$. 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

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

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

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

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

A.2 Proof of Lemma 1.4

Given Lemma 1.3, consider an arbitrary mediation plan $\mu^{\text{MB}} \in \Delta(Q \times \{r^{\text{MB}*}\})$ that satisfies (1.7) for all $q \in \text{supp}(\mu^{\text{MB}})$. 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 $x^{\text{MB}}_{\text{HH}}$ and $x^{\text{MB}}_{\text{LL}}$ 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:

$$x^{\text{MB}}_{\text{HH}} = \sum_{q:\, q(s_H,b_H)=p_B(b_H)} \mu^{\text{MB}}(q),$$
$$x^{\text{MB}}_{\text{LL}} = \sum_{q:\, q(s_L,b_L)=p_B(b_H)} \mu^{\text{MB}}(q).$$

Note that these probabilities are identical 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 $\mu^{\text{MB}*}$.[33]

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,

[33]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 optimal.

his expected payoff from misreporting under $\mu^{\mathrm{MB}*}$ is at most

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

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

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

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

High-type buyer. If the high-type buyer misreports his type under $\mu^{\mathrm{MB}*}$, 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 $\mu^{\mathrm{MB}*}$ is at most

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

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

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

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

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

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

which depends only on $x_{\mathrm{HH}}^{\mathrm{MB}}$ and $x_{\mathrm{LL}}^{\mathrm{MB}}$, implying that the expected social surplus is identical under both mediation plans. □

A.3 Proof of Proposition 1.2

I show that $x^{\mathrm{SO}}=(1,1)$ is the only point that 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

$$x_{\mathrm{LL}}^{\mathrm{SO}}\le 1-\frac{\pi_B(b_H-b_L)}{(1-\pi_B)(b_L-s_L)}\left(1-x_{\mathrm{LH}}^{\mathrm{SO}}\right). \tag{A.9}$$

In $x_{\mathrm{LH}}^{\mathrm{SO}}$-$x_{\mathrm{LL}}^{\mathrm{SO}}$ space, (A.9) 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 $x^{\mathrm{SO}}=(1,1)$.

□

A.4 Proof of Proposition 1.3

Solving the linear program

Assume $h > \bar{h}^{\text{Eff}}$ and C1–C4. Rearranging the constraints, (MB-P) can be written as follows:[34]

$$
\begin{aligned}
\max_{x \in \mathbb{R}^3} \quad & \pi_S \pi_B (b_H - s_H) x_{\text{HH}} + (1-\pi_S)\pi_B (b_H - s_L) \\
& \quad + (1-\pi_S)(1-\pi_B)(b_L - s_L)(1 - x_{\text{LL}}) \\
\text{subject to} \quad & g_1(x) = x_{\text{HH}} - x_{\text{LH}} \geq 0 \\
& g_2(x) = x_{\text{LL}} - x_{\text{LH}} \geq 0 \\
& g_3(x) = (1-\pi_B)(b_L - h s_L)(1 - x_{\text{LL}}) \\
& \qquad\qquad - \pi_B (b_H - b_L)(x_{\text{HH}} - x_{\text{LH}}) \geq 0 \\
& g_4(x) = \pi_S \pi_B b_H x_{\text{HH}} + (1-\pi_S)\pi_B (b_H - b_L) x_{\text{LH}} - (1-\pi_S)(1-\pi_B) b_L x_{\text{LL}} \geq 0 \\
& g_5(x) = -\pi_S \pi_B b_H (1 - x_{\text{HH}}) - (1-\pi_S)\pi_B (b_H - b_L)(1 - x_{\text{LH}}) \\
& \qquad\qquad + (1-\pi_S)(1-\pi_B) b_L (1 - x_{\text{LL}}) \geq 0 \\
& g_6(x) = 1 - x_{\text{HH}} \geq 0 \\
& g_7(x) = 1 - x_{\text{LH}} \geq 0 \\
& g_8(x) = 1 - x_{\text{LL}} \geq 0 \\
& g_9(x) = x_{\text{HH}} \geq 0 \\
& g_{10}(x) = x_{\text{LH}} \geq 0 \\
& g_{11}(x) = x_{\text{LL}} \geq 0.
\end{aligned}
$$

Let λ_i denote the Lagrange multiplier associated with g_i. By the KKT conditions, $x \in \mathbb{R}^3$ is the solution if and only if there exists $\lambda \in \mathbb{R}^{11}_+$ such that (x, λ) satisfies the following conditions:

Primal feasibility. $g_i(x) \geq 0$ for all $i \in \{1, \dots, 11\}$.

[34] For notational simplicity, the superscript "MB" is omitted in the proof.

Dual feasibility. $\lambda_i \geq 0$ for all $i \in \{1, \ldots, 11\}$, and

$$\pi_S \pi_B (b_H - s_H) + \lambda_1 - \pi_B (b_H - b_L)\lambda_3 + \pi_S \pi_B b_H (\lambda_4 + \lambda_5) - \lambda_6 + \lambda_9 = 0 \tag{A.10}$$

$$-\lambda_1 - \lambda_2 + \pi_B (b_H - b_L)\lambda_3 + (1 - \pi_S)\pi_B (b_H - b_L)(\lambda_4 + \lambda_5) - \lambda_7 + \lambda_{10} = 0 \tag{A.11}$$

$$\begin{aligned} &-(1 - \pi_S)(1 - \pi_B)(b_L - s_L) + \lambda_2 - (1 - \pi_B)(b_L - h s_L)\lambda_3 \\ &\quad - (1 - \pi_S)(1 - \pi_B) b_L (\lambda_4 + \lambda_5) - \lambda_8 + \lambda_{11} = 0. \end{aligned} \tag{A.12}$$

Complementary slackness. $\lambda_i g_i(x) = 0$ for all $i \in \{1, \ldots, 11\}$.

To simplify the exposition, I introduce the following notations:

$$\begin{aligned} x^*_{\text{HH}} &= \frac{(1 - \pi_B)(b_L - h s_L)}{\pi_B (b_H - b_L)}, \\ J &= \pi_S \pi_B b_H + (1 - \pi_S)\pi_B (b_H - b_L) - (1 - \pi_S)(1 - \pi_B) b_L, \\ K &= \pi_S \pi_B s_H (1 - x^*_{\text{HH}}) + (1 - \pi_S)\pi_B (b_H - b_L) - (1 - \pi_S)(1 - \pi_B) s_L, \end{aligned}$$

where x^*_{HH} is the value of x_{HH} at which $g_3(x) \geq 0$ binds when $x_{\text{LH}} = x_{\text{LL}} = 0$. Since $g_5(x) = g_4(x) - J$, and C3 $\Leftrightarrow$ $J \geq 0$, it follows that the intermediary follows the recommendation $p_B(b_H)$ whenever she follows the recommendation $p_B(b_L)$.

I first solve (MB-P) by considering three cases based on the value of J, and then derive the corresponding conditions on the ratio h that distinguish between them.

Case 1: When $J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + \min\{K, 0\}$

I show that the following (x, λ) satisfies the KKT conditions:

$$\begin{aligned} x^* &= \left(x^*_{\text{HH}}, 0, 0\right), \\ \lambda^*_2 &= \pi_S \pi_B (b_H - s_H), \\ \lambda^*_3 &= \frac{\pi_S (b_H - s_H)}{b_H - b_L}, \\ \lambda^*_{11} &= (1 - \pi_S)(1 - \pi_B)(b_L - s_L) - \pi_S \pi_B (b_H - s_H)(1 - x^*_{\text{HH}}), \\ \lambda^*_i &= 0 \;\text{ for all } i \notin \{2, 3, 11\}. \end{aligned}$$

Primal feasibility. It is easy to see $g_i(x^*) \geq 0$ for all $i \notin \{5, 6\}$. Note that $g_5(x^*) \geq 0$ is reduced to $J \leq \pi_S \pi_B b_H x^*_{\text{HH}}$ and that $g_6(x^*) = 1 - x^*_{\text{HH}} \geq 0 \Leftrightarrow h \geq \bar{h}^{\text{Eff}}$.

Dual feasibility. λ^*_2 and λ^*_3 are positive, and

$$\lambda^*_{11} \geq 0 \iff J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + K.$$

The left-hand side of (A.10) is reduced to

$$\pi_S \pi_B (b_H - s_H) - \pi_S \pi_B (b_H - s_H) = 0.$$

Next, (A.11) is satisfied since $\lambda^*_2 = \pi_B (b_H - b_L) \lambda^*_3$. Finally, the left-hand side of (A.12) is reduced to

$$\pi_S \pi_B (b_H - s_H) x^*_{\text{HH}} - \underbrace{(1 - \pi_B)(b_L - h s_L)}_{= \pi_B (b_H - b_L) x^*_{\text{HH}}} \frac{\pi_S (b_H - s_H)}{b_H - b_L} = 0.$$

Complementary slackness. The conditions are satisfied since $\lambda^*_i = 0$ for all $i \notin \{2, 3, 11\}$ and $g_2(x^*) = g_3(x^*) = g_{11}(x^*) = 0$.

Case 2: When $\pi_S \pi_B b_H x^*_{\text{HH}} < J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + K$

I show that the following (x, λ) satisfies the KKT conditions:

$$\begin{aligned}
x^* &= (1, 1, 1), \\
\lambda^*_2 &= \pi_S \pi_B (b_H - s_H) + [\pi_S \pi_B b_H + (1 - \pi_S) \pi_B (b_H - b_L)] \lambda^*_5, \\
\lambda^*_3 &= \frac{\pi_S \pi_B (b_H - s_H) + \pi_S \pi_B b_H \lambda^*_5}{\pi_B (b_H - b_L)}, \\
\lambda^*_5 &= \frac{(1 - \pi_S)(1 - \pi_B)(b_L - s_L) - \pi_S \pi_B (b_H - s_H)(1 - x^*_{\text{HH}})}{J - \pi_S \pi_B b_H x^*_{\text{HH}}}, \\
\lambda^*_i &= 0 \ \text{ for all } i \notin \{2, 3, 5\}.
\end{aligned}$$

Primal feasibility. It is easy to see $g_i(x^*) \geq 0$ for all $i \neq 4$ and $g_4(x^*) = J > 0$ by assumption.

Dual feasibility. Note that λ_2^* and λ_3^* are positive if λ_5^* is nonnegative. Since the nominator of λ_5^* equals λ_{11}^* of Case 1, $\lambda_5^* \geq 0$ by assumption. The left-hand side of (A.10) is reduced to

$$\pi_S \pi_B (b_H - s_H) - \pi_S \pi_B (b_H - s_H) - \pi_S \pi_B b_H \lambda_5^* + \pi_S \pi_B b_H \lambda_5^* = 0.$$

Next, the left-hand side of (A.11) is reduced to

$$\begin{aligned} &- \pi_S \pi_B (b_H - s_H) - [\pi_S \pi_B b_H + (1 - \pi_S) \pi_B (b_H - b_L)] \lambda_5^* \\ &\quad + \pi_S \pi_B (b_H - s_H) + \pi_S \pi_B b_H \lambda_5^* + (1 - \pi_S) \pi_B (b_H - b_L) \lambda_5^* = 0. \end{aligned}$$

Finally, the left-hand side of (A.12) is reduced to

$$\begin{aligned} &- (1 - \pi_S)(1 - \pi_B)(b_L - s_L) + \pi_S \pi_B (b_H - s_H) + [\pi_S \pi_B b_H + (1 - \pi_S) \pi_B (b_H - b_L)] \lambda_5^* \\ &\quad - \underbrace{(1 - \pi_B)(b_L - h s_L)}_{= \pi_B (b_H - b_L) x_{\text{HH}}^*} \frac{\pi_S \pi_B (b_H - s_H) + \pi_S \pi_B b_H \lambda_5^*}{\pi_B (b_H - b_L)} - (1 - \pi_S)(1 - \pi_B) b_L \lambda_5^* \\ &= \pi_S \pi_B (b_H - s_H)(1 - x_{\text{HH}}^*) - (1 - \pi_S)(1 - \pi_B)(b_L - s_L) + \left(J - \pi_s \pi_B b_H x_{\text{HH}}^*\right) \lambda_5^* \\ &= 0. \end{aligned}$$

Complementary slackness. The conditions are satisfied since $\lambda_i^* = 0$ for all $i \notin \{2, 3, 5\}$ and $g_2(x^*) = g_3(x^*) = g_5(x^*) = 0$.

Case 3: When $J > \pi_S \pi_B b_H x^*_{\text{HH}} + K$

I show that the following (x, λ) satisfies the KKT conditions:

$$\begin{aligned}
x^* &= (1, 1, 1) \\
\lambda_2^* &= \pi_S \pi_B (b_H - s_H), \\
\lambda_3^* &= \frac{\pi_S (b_H - s_H)}{b_H - b_L}, \\
\lambda_8^* &= \pi_S \pi_B (b_H - s_H)(1 - x^*_{\text{HH}}) - (1 - \pi_S)(1 - \pi_B)(b_L - s_L), \\
\lambda_i^* &= 0 \text{ for all } i \notin \{2, 3, 8\}.
\end{aligned}$$

Primal feasibility. It is easy to see $g_i(x^*) \geq 0$ for all $i \neq 4$ and $g_4(x^*) = J \geq 0$ by assumption.

Dual feasibility. Note that λ_2^* and λ_3^* coincide with those of Case 1 and that $-\lambda_8^*$ equals λ_{11}^* of Case 1. Hence, (A.10)–(A.12) are satisfied, and λ_8^* is positive by assumption.

Complementary slackness. The conditions are satisfied since $\lambda_i^* = 0$ for all $i \notin \{2, 3, 8\}$ and $g_2(x^*) = g_3(x^*) = g_8(x^*) = 0$.

Hence, in all three cases, the pair (x^*, λ^*) satisfies the KKT conditions. To summarize, under $h > \bar{h}^{\text{Eff}}$ and C1–C4, the solution to (MB-P) is

- $x^* = \left(x^*_{\text{HH}}, 0, 0\right)$ if $J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + \min\{K, 0\}$;
- $x^* = (1, 1, 1)$ otherwise.

Deriving the Conditions on the Ratio h

Define J_s as follows:

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

Note that

$$
\begin{aligned}
&J \leq \pi_S \pi_B b_H x^*_{HH} \iff h \leq \frac{b_L}{s_L} - \frac{(b_H - b_L)J}{\pi_S(1-\pi_B)b_H s_L} \equiv h_J,\\
&J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + K \iff h \leq \frac{b_L}{s_L} - \frac{(b_H - b_L)(J - J_s)}{\pi_S(1-\pi_B)(b_H - s_H)s_L} \equiv h_{J-K},\\
&K \leq 0 \iff h \leq \frac{b_L}{s_L} - \frac{(b_H - b_L)J_s}{\pi_S(1-\pi_B)s_H s_L} \equiv h_K,\\
&J \geq 0 \iff \pi_B \geq \frac{(1-\pi_S)b_L}{b_H} \equiv \pi_B^J(\pi_s),\\
&J_s \geq 0 \iff \pi_B \geq \frac{(1-\pi_S)s_L}{\pi_S s_H + (1-\pi_S)(b_H - b_L + s_L)} \equiv \pi_B^{J_s}(\pi_s),\\
&J - J_s \geq 0 \iff \pi_B \geq \frac{(1-\pi_S)(b_L - s_L)}{\pi_S(b_H - s_H) + (1-\pi_S)(b_L - s_L)} \equiv \pi_B^{J-J_s}(\pi_s).
\end{aligned}
$$

Since $J \leq \pi_S \pi_B b_H x^*_{\text{HH}} + \min\{K, 0\}$ is reduced to $h \leq h_J$ when $K \geq 0$ and to $h \leq h_{J-K}$ when $K < 0$, it is instructive to identify the conditions under which h_J and h_{J-K} respectively exceed one. As briefly discussed in the paragraph preceding Proposition 1.3, h_J and h_{J-K} exceed one if and only if π_B is smaller than the respective thresholds $\pi_B^{h_J}$ and $\pi_B^{h_{J-K}}$:

$$
\begin{aligned}
h_J > 1 &\iff \pi_B < \pi_B^{h_J}(\pi_S),\\
h_{J-K} > 1 &\iff \pi_B < \pi_B^{h_{J-K}}(\pi_S).
\end{aligned}
$$

It can also be verified that

$$
h_{J-K} \leq h_J \leq h_K \iff \pi_B \leq \frac{b_H s_L - s_H b_L}{b_H(b_H - s_H - b_L + s_L)} \equiv \pi_B^\star
$$

and that C1 and C2 imply $\pi_B^\star > 0$. Recall that C3 $\Leftrightarrow \pi_B \geq \pi_B^J(\pi_s) \Leftrightarrow J \geq 0$. To summarize, four possible regions arise, as shown in Figure A.1:

1a. If (π_S, π_B) lies in Region 1a, then $J_s \geq 0$, $J - J_s \geq 0$, and $\pi_B \geq \pi_B^\star$, which together imply $h_K \leq h_J \leq h_{J-K} \leq z\frac{b_L}{s_L}$. Since $h_J > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_J$ and $(1,1,1)$ otherwise.

1b. If (π_S, π_B) lies in Region 1b, then $J_s \geq 0$, $J - J_s \geq 0$, and $\pi_B < \pi_B^\star$, which together imply $h_{J-K} < h_J < h_K \leq \frac{b_L}{s_L}$. Since $h_{J-K} > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_{J-K}$

and $(1, 1, 1)$ otherwise.

2. If (π_S, π_B) lies in Region 2, then $J_s \geq 0$, $J - J_s < 0$, and $\pi_B \geq \pi_B^\star$, which together imply $h_K \leq h_J \leq \frac{b_L}{s_L} < h_{J-K}$. Again, since $h_J > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_J$ and $(1, 1, 1)$ otherwise.

3. If (π_S, π_B) lies in Region 3, then $J_s < 0$, $J - J_s \geq 0$, and $\pi_B < \pi_B^\star$, which together imply $h_{J-K} < h_J \leq \frac{b_L}{s_L} < h_K$. Again, since $h_{J-K} > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_{J-K}$ and $(1, 1, 1)$ otherwise.

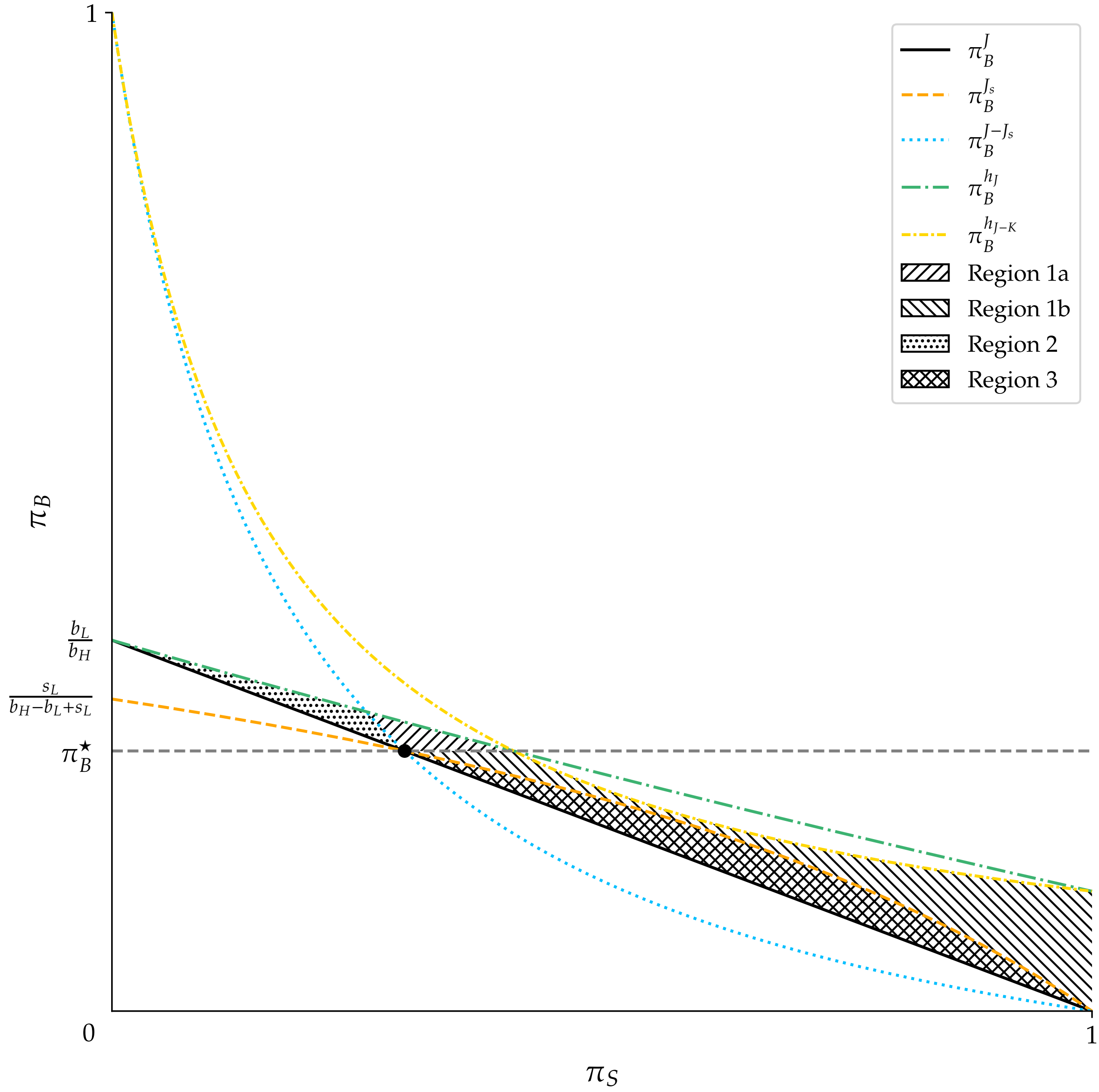


Figure A.1: Four possible regions under C1–C4. Parameter values are $b_H = 3.5$, $s_H = 1.5$, $b_L = 1.3$, and $s_L = 1.0$. Note that $h_J, h_{J-K} \leq 1$ when $\pi_B \geq \min\left\{\pi_B^{h_J}, \pi_B^{h_{J-K}}\right\}$, so the solution is $(1, 1, 1)$ for all $h \leq \frac{b_L}{s_L}$.

Ex Ante Expected Social Surplus

If $h \leq h^*$, the SB outcome is characterized by the mediation plan $\left(x_{\mathrm{HH}}^{\mathrm{MB}*}, 0, 0\right)$. Under this mediation plan, the intermediary offers $p_B(b_H)$ for (s_H, b_H) with probability x_{HH}^* and $p_B(b_L)$ with probability $1 - x_{\mathrm{HH}}^*$. For (s_L, b_H) and (s_L, b_L), she offers $p_B(b_L)$ with probability one. As a result, the traders (s_H, b_H) trade at the price $p_B(b_H)$ with probability x_{HH}^*, while the traders (s_L, b_H) and (s_L, b_L) trade at the price $p_B(b_L)$ with probability one. The associated ex ante expected social surplus is

$$\pi_S \pi_B (b_H - s_H) x_{\mathrm{HH}}^* + (1 - \pi_S)\pi_B(b_H - s_L) + (1 - \pi_S)(1 - \pi_B)(b_L - s_L). \tag{A.13}$$

Since x_{HH}^* is decreasing in h, so is the expected social surplus.

If $h > h^*$, the SB outcome is characterized by the mediation plan $(1, 1, 1)$. Under this mediation plan, the intermediary always offers $p_B(b_H)$ whenever there is a gain from trade. Thus, the traders (s_H, b_H) and (s_L, b_H) trade at the price $p_B(b_H)$ with probability one, while (s_L, b_L) do not trade with probability one. The associated ex ante expected social surplus is

$$\pi_S \pi_B (b_H - s_H) + (1 - \pi_S)\pi_B(b_H - s_L), \tag{A.14}$$

which is independent of h.

The difference between (A.13) and (A.14) is

$$\begin{aligned} & - \pi_S \pi_B (b_H - s_H)(1 - x_{\mathrm{HH}}^*) + (1 - \pi_S)(1 - \pi_B)(b_L - s_L) \\ &= \pi_S \pi_B b_H x_{\mathrm{HH}}^* + K - J, \end{aligned}$$

which is nonnegative because $h \leq h^* \Rightarrow J \leq \pi_S \pi_B b_H x_{\mathrm{HH}}^* + K$. Hence, the expected social surplus is weakly higher under $\left(x_{\mathrm{HH}}^*, 0, 0\right)$.

A.5 Second-Best Outcome When Any of C1–C3 Fails

First, assume that C1, C2, and C4 hold but C3 does not. That is, consider the alternative condition:

C3′. $\pi_B < \frac{(1-\pi_S)b_L}{b_H}$.

Under C1, C2, C3′, and C4, four possible regions arise, as shown in Figure A.2:

4. If (π_S, π_B) lies in Region 4, then $J < 0$, $J_s \geq 0$, $J - J_s < 0$, and $\pi_B \geq \pi_B^\star$, which together imply $h_K \leq \frac{b_L}{s_L} < h_J \leq h_{J-K}$. The solution is $\left(x^*_{\mathrm{HH}}, 0, 0\right)$ for all $h \leq \frac{b_L}{s_L}$.

5. If (π_S, π_B) lies in Region 5, then $J < 0$, $J_s < 0$, $J - J_s \geq 0$, and $\pi_B < \pi_B^\star$, which together imply $h_{J-K} \leq \frac{b_L}{s_L} < h_J < h_K$. Since $h_{J-K} > 1$, the solution is $\left(x^*_{\mathrm{HH}}, 0, 0\right)$ if $h \leq h_{J-K}$ and $\left(\hat{x}^*_{\mathrm{HH}}, x^*_{\mathrm{L}}, x^*_{\mathrm{L}}\right)$ otherwise, where

$$\hat{x}^*_{\mathrm{HH}} = x^*_{\mathrm{HH}} + \left(1 - x^*_{\mathrm{HH}}\right)x^*_{\mathrm{L}},$$
$$x^*_{\mathrm{L}} = \frac{\pi_S \pi_B b_H x^*_{\mathrm{HH}}}{\pi_S \pi_B b_H x^*_{\mathrm{HH}} - J}.$$

This case is discussed in detail below.

6a. If (π_S, π_B) lies in Region 6a, then $J < 0$, $J_s < 0$, $J - J_s < 0$, and $\pi_B \geq \pi_B^\star$, which together imply $\frac{b_L}{s_L} < h_K \leq h_J \leq h_{J-K}$. The solution is $\left(x^*_{\mathrm{HH}}, 0, 0\right)$ for all $h \leq \frac{b_L}{s_L}$.

6b. If (π_S, π_B) lies in Region 6b, then $J < 0$, $J_s < 0$, $J - J_s < 0$, and $\pi_B < \pi_B^\star$, which together imply $\frac{b_L}{s_L} < h_{J-K} < h_J < h_K$. The solution is again $\left(x^*_{\mathrm{HH}}, 0, 0\right)$ for all $h \leq \frac{b_L}{s_L}$.

In Case 5, the mediation plan $(1, 1, 1)$ does not satisfy the KKT conditions because the primal feasibility requires $J \geq 0$. In thi case, $\left(\hat{x}^*_{\mathrm{HH}}, x^*_{\mathrm{L}}, x^*_{\mathrm{L}}\right)$ is the solution when $h > h_{J-K}$. Note that $x^*_{\mathrm{HH}} + \left(1 - x^*_{\mathrm{HH}}\right)x_{\mathrm{L}}$ is the value of x_{HH} at which (MB-IC$_{s_L}$) binds when $x_{\mathrm{LH}} = x_{\mathrm{LL}} = x_{\mathrm{L}}$ and that x^*_{L} is the value of x_{L} at which (IC$_{\mathrm{Int}}$-1) binds when $x_{\mathrm{HH}} = x^*_{\mathrm{HH}} + \left(1 - x^*_{\mathrm{HH}}\right)x_{\mathrm{L}}$.

I show that the following λ satisfies the KKT conditions along with $\left(\hat{x}^*_{\mathrm{HH}}, x^*_{\mathrm{L}}, x^*_{\mathrm{L}}\right)$

when $h_{J-K} < h \leq \frac{b_L}{s_L} \Leftrightarrow \pi_S \pi_B b_H x^*_{\text{HH}} + K < J < \pi_S \pi_B b_H x^*_{\text{HH}}$:

$$\lambda^*_2 = \pi_B(b_H - b_L)\big[\lambda^*_3 + (1-\pi_S)\lambda^*_4\big],$$
$$\lambda^*_3 = \frac{\pi_S(b_H - s_H) + \pi_S b_H \lambda^*_4}{b_H - b_L},$$
$$\lambda^*_4 = \frac{J - \pi_S \pi_B b_H x^*_{\text{HH}} - K}{\pi_S \pi_B b_H x^*_{\text{HH}} - J},$$
$$\lambda^*_i = 0 \ \text{ for all } i \notin \{2,3,4\}.$$

Primal feasibility. It is easy to see $g_i(x^*) \geq 0$ for all $i \notin \{5,6\}$. Note that $g_5(x^*) \geq 0$ is reduced to $J \leq \pi_S \pi_B b_H x^*_{\text{HH}}$ and that $g_6(x^*) = 1 - x^*_{\text{HH}} \geq 0 \Leftrightarrow h \geq \bar{h}^{\text{Eff}}$.

Dual feasibility. λ^*_4 is nonnegative by assumption. Hence, λ^*_2 and λ^*_3 are also nonnegative. The left-hand side of (A.10) is reduced to

$$\pi_S \pi_B(b_H - s_H) - \pi_S \pi_B(b_H - s_H) - \pi_S \pi_B b_H \lambda^*_4 + \pi_S \pi_B b_H \lambda^*_4 = 0.$$

The left-hand side of (A.11) is reduced to

$$-\pi_B(b_H - b_L)\big[\lambda^*_3 + (1-\pi_S)\lambda^*_4\big] + \pi_B(b_H - b_L)\lambda^*_3 + (1-\pi_S)\pi_B(b_H - b_L)\lambda^*_4 = 0.$$

The left-hand side of (A.12) is reduced to

$$\begin{aligned}
&-(1-\pi_S)(1-\pi_B)(b_L - s_L) + \pi_B(b_H - b_L)(1 - x^*_{\text{HH}})\lambda^*_3 \\
&\quad + [(1-\pi_S)\pi_B(b_H - b_L) - (1-\pi_S)(1-\pi_B)b_L]\lambda^*_4 \\
&= \underbrace{\pi_S \pi_B (1 - x^*_{\text{HH}})(b_H - s_H) - (1-\pi_S)(1-\pi_B)(b_L - s_L)}_{= J - \pi_S \pi_B b_H x^*_{\text{HH}} - K} \\
&\quad + \underbrace{\big[\pi_S \pi_B b_H (1 - x^*_{\text{HH}}) + (1-\pi_S)\pi_B(b_H - b_L) - (1-\pi_S)(1-\pi_B)b_L\big]}_{= J - \pi_S \pi_B b_H x^*_{\text{HH}}} \lambda^*_4 \\
&= 0.
\end{aligned}$$

Complementary slackness. The conditions are satisfied since $\lambda_i^* = 0$ for all $i \notin \{2, 3, 4\}$ and $g_2(x^*) = g_3(x^*) = g_4(x^*) = 0$.

Hence, (x^*, λ^*) satisfies the KKT conditions.

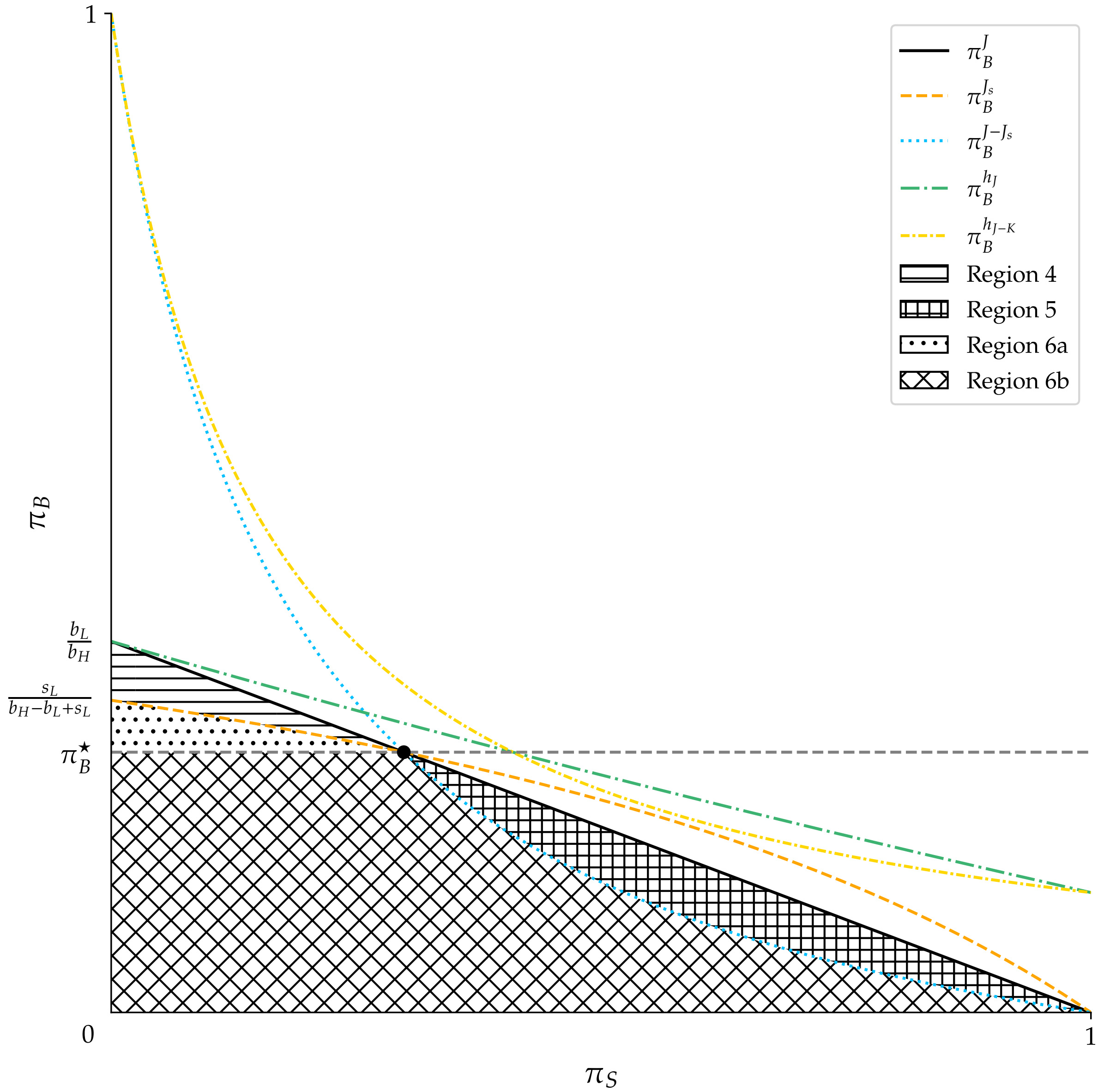


Figure A.2: Four possible regions under C1, C2, C3′, and C4. Parameter values are $b_H = 3.5$, $s_H = 1.5$, $b_L = 1.3$, and $s_L = 1.0$.

Next, assume that C1 and C4 hold but C2 does not. That is, consider the alternative condition:

C2′. $\frac{b_L}{s_L} \geq \frac{b_H}{s_H}$.

Under C1, C2′, and C4, I have $\pi_B^\star \leq 0$, which implies $h_K < h_J < h_{J-K}$ for all $\pi_B \in (0,1)$. In this case, four possible regions arise, as shown in Figure A.3:

1a. If (π_S, π_B) lies in Region 1a, then $J \geq 0$, $J_s \geq 0$, and $J - J_s \geq 0$, which together imply $h_K < h_J < h_{J-K} \leq \frac{b_L}{s_L}$. Since $h_J > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq \min\left\{h_J, \frac{b_H}{s_H}\right\}$ and $(1,1,1)$ otherwise.

2. If (π_S, π_B) lies in Region 2, then $J \geq 0$, $J_s \geq 0$ and $J - J_s < 0$, which together imply $h_K < h_J \leq \frac{b_L}{s_L} < h_{J-K}$. Again, since $h_J > 1$, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq \min\left\{h_J, \frac{b_H}{s_H}\right\}$ and $(1,1,1)$ otherwise.

4. If (π_S, π_B) lies in Region 4, then $J < 0$, $J_s \geq 0$, and $J - J_s < 0$, which together imply $h_K \leq \frac{b_L}{s_L} < h_J < h_{J-K}$. The solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ for all $h \leq \frac{b_H}{s_H}$.

6a. If (π_S, π_B) lies in Region 6a, then $J < 0$, $J_s < 0$, and $J - J_s < 0$, which together imply $\frac{b_L}{s_L} < h_K < h_J < h_{J-K}$. The solution is again $\left(x^*_{\text{HH}}, 0, 0\right)$ for all $h \leq \frac{b_H}{s_H}$.

Note that

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

Regions 1a and 2 are divided by the curve defined by the right-hand side of this inequality. Above it, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_J$ and $(1,1,1)$ otherwise. Below it, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ for all $h \leq \frac{b_H}{s_H}$.

Finally, if C1 does not hold, then this implies C2′ with a strict inequality, which in turn implies $h_{J-K} \leq h_J \leq h_K \Leftrightarrow \pi_B \geq \pi_B^\star > 1$; thus, $h_K < h_J < h_{J-K}$ for all $\pi_B \in (0,1)$.[35] Since $\pi_B^{J-J_s}(\pi_S) > \pi_B^J(\pi_S) > \pi_B^{J_s}(\pi_S)$ for all $\pi_S \in (0,1)$ as in Figure A.3, this case is essentially the same as the previous case, where C1, C2′, and C4 hold.

[35]Note that $b_L - s_L \geq b_H - s_H$ implies $b_H(b_L - s_L) > b_L(b_H - s_H)$, which is equivalent to $\frac{b_L}{s_L} > \frac{b_H}{s_H}$.

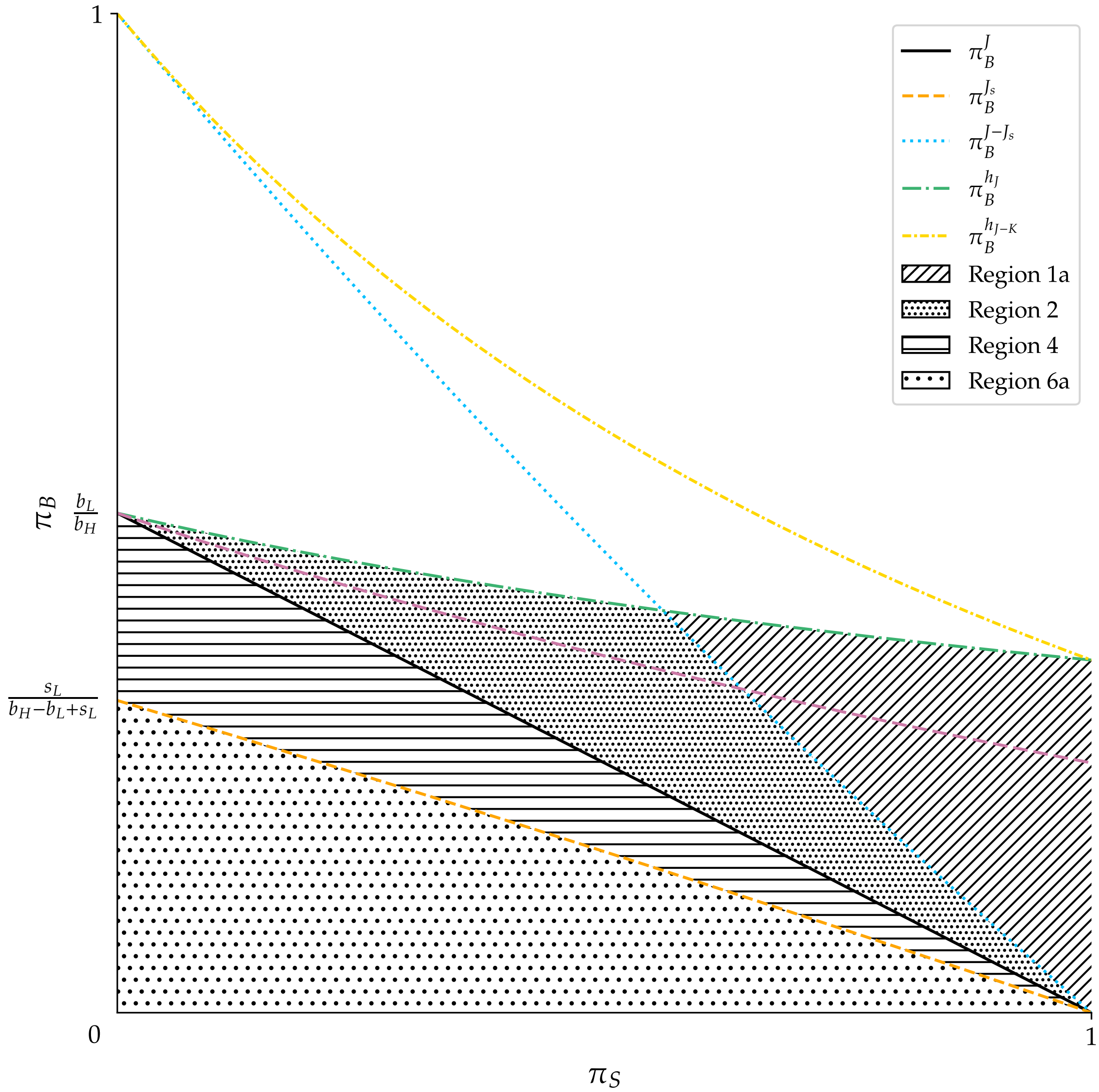


Figure A.3: Four possible regions under C1, C2′, and C4. Parameter values are $b_H = 2.2$, $s_H = 1.5$, $b_L = 1.1$, and $s_L = 0.5$. Note that $h_J \leq 1$ when $\pi_B \geq \pi_B^{h_J}$, so the solution is $(1, 1, 1)$ for all $h \leq \frac{b_H}{s_H}$. The magenta curve shows the threshold where $h_J = \frac{b_H}{s_H}$: above this curve, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ if $h \leq h_J$ and $(1, 1, 1)$ otherwise; below it, the solution is $\left(x^*_{\text{HH}}, 0, 0\right)$ for all $h \leq \frac{b_H}{s_H}$. This curve is not included in the legend to avoid visual clutter. This figure also applies to the case where C1 fails.

Chapter 2

Moral Hazard with Dual Risk-Averse Agent

2.1 Introduction

Since the seminal work of Holmström (1979), an extensive literature on moral hazard has studied how a principal can incentivize an agent to take desirable actions when the agent's action is not contractible. A persistent critique of the theory is that it often predicts optimal contracts that are complex and highly sensitive to the underlying distribution of outcomes, in contrast to the simple contracts frequently observed in practice.

This chapter offers a new explanation for linear contracts by showing that *dual risk-aversion* implies the optimality of debt contracts. I consider a stylized moral hazard model with a single agent who is protected by the limited liability constraint and *dual risk-averse* in the sense of Yaari (1987); that is, he overweights the probabilities of "bad" outcomes when making decisions. I assume a specific form of dual risk-aversion where, if outcome follows distribution F, the agent behaves as if he believes that the worst outcome is realized with probability $(1 - k)$ and that outcome follows F with probability k, where $k \in (0, 1)$. I restrict the class of admissible contracts to those nondecreasing and with slope at most 1. To characterize the principal's cost-minimizing contract for inducing each action, I first consider the *relaxed* problem, where the agent's incentive compatibility constraint is replaced by its first-order condition, the so-called *FOA approach*. The relaxed problem is reduced to an optimal control problem, and Pontryagin's maximum principle characterizes its solution. I show that, under the strict version of the monotone likelihood ratio property, the solution to the relaxed problem is *debt contract*; that is, wage is constant up to some level of outcome and grows at the same rate as outcome afterward (Proposition 2.1). I also show that this contract is indeed the solution to the original problem—the FOA approach is valid—if the outcome distribution is (in proportionate terms) less convex in action than the disutility of action (Proposition 2.2).

This chapter contributes to the literature on moral hazard in two respects. First, I offer a new explanation, dual risk-aversion, for linear contracts observed in practice. Proposition 2.1 shows the optimality of debt contracts under the monotonicity and the limited liability constraints, which resonantes with the finding of Innes (1990) and Poblete and Spulber (2012), who show the optimality of debt contracts under similar constraints.[1] Relatedly, Holmström and Milgrom (1987) and Carroll (2015) show the optimality of linear contracts in certain environments. Second, I derive a sufficient condition for validating the FOA. The seminal work of Rogerson (1985) shows that the FOA is valid if the distribution of outcome is convex in action (Convexity of Distribution Function Condition).[2] Recently, Chade and Swinkels (2020) establish a more general sufficient condition for validating the FOA. Leveraging their technique, I derive a sufficient condition in my environment (Proposition 2.2). As is the case in Chade and Swinkels (2020), this condition is implied by Convexity of Distribution Function Condition and convex disutility.

As an application of dual theory, this chapter is also related to the set of papers, Gershkov, Moldovanu, Strack and Zhang (2021, 2023, 2025). Except Gershkov et al. (2021), they show that dual risk-aversion can better explain observed consumer behavior

[1] Poblete and Spulber (2012) also show that a capped bonus contract, another form of linear contract, is optimal when the "critical ratio" is decreasing in the state.

[2] Since Convexity of Distribution Function Condition is restrictive, Jewitt (1988) establishes an alternative condition and generalizes it to the case with more than one observable variable.

and forms of contracts than expected utility theory, which resonates with the message of this chapter.

The remainder of the chapter is organized as follows. Section 2.2 introduces the model. I characterize the solution to the relaxed problem in Section 2.3 and derive a sufficient condition for validating the FOA in Section 2.4. In Section 2.5, I discuss the implications of some modeling assumptions. Section 2.6 concludes the chapter.

2.2 Model

Environment: I consider a stylized moral hazard environment with a single principal and a single agent.[3] The agent takes a hidden action $a \in [0, \overline{a}]$ at a disutility $\psi(a)$, which is strictly increasing and C^2. The disutility ψ is normalized so that $\psi(0) = 0$. To avoid corner solutions, I further assume the Inada conditions: $\psi'(0) = 0$ and $\psi'(\overline{a}) = +\infty$. The agent's action affects the distribution of outcome x, which takes a value in $\left[\underline{x}, \overline{x}\right]$. Given action a, the outcome x has a distribution $F(x \mid a)$ with everywhere positive density $f(x \mid a)$. I assume that F and f are C^2 and that the agent's action produces a first-order stochastic dominant shift on the outcome; that is, $F_a(x \mid a) < 0$ for all $x \in \left(\underline{x}, \overline{x}\right)$ and all a.[4]

Preferences: The principal offers a bounded wage schedule $w\colon \left[\underline{x}, \overline{x}\right] \to [-m, M]$ to the agent, where $m, M \in \mathbb{R}_+$ and M is sufficiently large. She is risk-neutral and maximizes her expected net benefit $\mathbb{E}[x - w(x)]$. The agent is *dual risk-averse* in the sense of Yaari (1987): given the distribution of wage $\Gamma^w(\cdot \mid a)$ induced by wage schedule w and action a, his gross utility is given by

$$\int_{-m}^{M} \phi(1 - \Gamma^w(v \mid a))\, \mathrm{d}v, \tag{2.1}$$

[3]I use feminine pronouns for the principal and masculine pronouns for the agent.
[4]Note that $F_a\left(\underline{x} \mid a\right) = F_a(\overline{x} \mid a) = 0$ for all a since the support of the outcome is fixed.

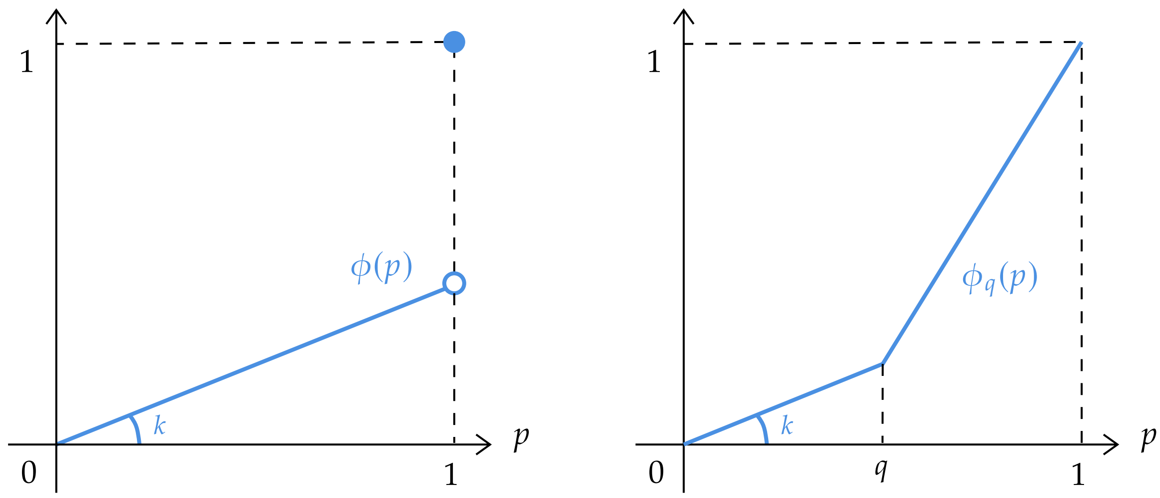


Figure 2.1: Distortion function ϕ can be interpreted as the limit of ϕ_q as q goes to 1.

where distortion function $\phi\colon [0,1] \to [0,1]$ is defined as

$$\phi(p) = \begin{cases} kp & \text{if } 0 \le p < 1 \\ 1 & \text{if } p = 1 \end{cases}$$

for some $k \in (0,1)$.

Remark 2.1. Yaari (1987) characterizes the dual representation for general continuous and nondecreasing ϕ satisfying $\phi(0) = 0$ and $\phi(1) = 1$, and risk-aversion is characterized by the convexity of ϕ. The present ϕ can be interpreted as the limit of the convex and piecewise-linear function that has a slope k on $[0,q)$ and extends from (q,kq) to $(1,1)$. See Figure 2.1.

Admissible wage schedules: I assume $m = 0$ so that the agent is protected by the limited liability constraint. Following Poblete and Spulber (2012), I restrict attention to wage schedules that satisfy the following assumptions.

Assumption 2.1 (Monotonicity)**.** $w(x)$ is nondecreasing.

Assumption 2.2 (Bounded Slope)**.** $x - w(x)$ is nondecreasing.

As discussed in Poblete and Spulber (2012), Assumption 2.1 prevents the agent from sabotaging the task while Assumption 2.2 prevents the principal from doing so after the action has been chosen. The latter also prevents the agent from inflating the outcome by borrowing money to increase his returns. For $x \geq x'$, Assumptions 2.1 and 2.2 imply $w(x) \geq w(x')$ and $x - w(x) \geq x' - w(x')$, respectively. By taking the limit, this implies that w is almost everywhere differentiable and Lipschitz continuous with

$$0 \leq w'(x) \leq 1 \quad \text{for almost every } x. \tag{Lipschitz}$$

2.3 Principal's (relaxed) problem

Suppose that the principal wants to induce action $a > 0$. Given Assumption 2.1, the distribution of wage is given by $\Gamma^w(v \mid a) = F\big(w^{-1}(v) \mid a\big)$, where w^{-1} is the generalized inverse of w:

$$w^{-1}(v) = \sup\{x \colon w(x) \leq v\}.$$

The agent's gross utility (2.1) can be written as

$$\begin{aligned}\int_0^M \phi\big(1 - F\big(w^{-1}(v) \,\big|\, a\big)\big)\,\mathrm{d}v &= w\big(\underline{x}\big) + \int_{\underline{x}}^{\overline{x}} \phi(1 - F(x \mid a))\, w'(x)\,\mathrm{d}x \\ &= w\big(\underline{x}\big) + \int_{\underline{x}}^{\overline{x}} k[1 - F(x \mid a)]\, w'(x)\,\mathrm{d}x,\end{aligned}$$

where the first equality follows from the change of variable $x = w^{-1}(v)$. Integrating by parts, this can be further transformed into

$$(1-k)w\big(\underline{x}\big) + k\int_{\underline{x}}^{\overline{x}} w(x) f(x \mid a)\,\mathrm{d}x. \tag{2.2}$$

That is, the agent behaves as if he believes that the worst outcome $\underline{x}$ is realized with probability $1 - k$ and that outcome follows F with probability k. By doing so, the agent overweights the probabilities of bad outcomes in making his decision, which is consistent with Yaari (1987)'s interpretation of dual risk-aversion.

For the agent to accept wage schedule w and choose action a, his net utility $U(w, a)$

must be greater than or equal to his reservation utility U_0:

$$U(w,a) \equiv w(\underline{x}) + \int_{\underline{x}}^{\overline{x}} k[1 - F(x \mid a)]\, w'(x)\, \mathrm{d}x - \psi(a) \geq U_0. \tag{PC}$$

Given w, he chooses action a if it maximizes $U(w, \tilde{a})$:

$$U(w,a) \geq U(w,\tilde{a}) \quad \text{for all } \tilde{a} \in [0, \overline{a}]. \tag{IC}$$

Its first-order condition $U_a(w,a) = 0$ is equivalent to

$$\int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)] w'(x)\, \mathrm{d}x = \psi'(a). \tag{FOA}$$

Since w is nondecreasing by (Lipschitz), the limited liability constraint is reduced to

$$w(\underline{x}) \geq 0. \tag{LL}$$

If the principal induces action a, she minimizes the expected payment $\int_{\underline{x}}^{\overline{x}} w(x) f(x \mid a)\, \mathrm{d}x$ to the agent under the constraints (PC), (IC), (LL), and (Lipschitz). To characterize the solution to this problem, I first consider a relaxed problem where (IC) is replaced by (FOA). Then, I derive a sufficient condition under which the solution to the relaxed problem indeed satisfies (IC); that is, the FOA is valid. The relaxed problem can be formulated as follows:

$$\min \quad \left[\int_{\underline{x}}^{\overline{x}} w(x) f(x \mid a)\, \mathrm{d}x \right]$$

$$\text{subject to} \quad w(\underline{x}) + \int_{\underline{x}}^{\overline{x}} k[1 - F(x \mid a)]\, w'(x)\, \mathrm{d}x - \psi(a) \geq U_0 \tag{PC}$$

$$\int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)] w'(x)\, \mathrm{d}x = \psi'(a) \tag{FOA}$$

$$w(\underline{x}) \geq 0 \tag{LL}$$

$$0 \leq w'(x) \leq 1 \qquad \text{for a.e. } x \tag{Lipschitz}$$

Given (Lipschitz), (FOA) can be satisfied for a only if

$$\int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)]\,\mathrm{d}x \geq \psi'(a). \tag{2.3}$$

In the remainder of the chapter, I restrict attention to those actions satisfying (2.3) with a strict inequality:[5]

$$A^* = \left\{ a \in (0, \overline{a}) \colon \int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)]\,\mathrm{d}x > \psi'(a) \right\}$$

To make the problem amenable to the control-theoretic technique, define α and β by

$$\alpha(x) = w(\underline{x}) + \int_{\underline{x}}^{x} k[1 - F(\tilde{x} \mid a)]\,w'(\tilde{x})\,\mathrm{d}\tilde{x}, \quad \alpha'(x) = k[1 - F(x \mid a)]\,w'(x),$$

$$\beta(x) = \int_{\underline{x}}^{x} k[-F_a(\tilde{x} \mid a)]w'(\tilde{x})\,\mathrm{d}\tilde{x}, \qquad \beta'(x) = k[-F_a(x \mid a)]\,w'(x).$$

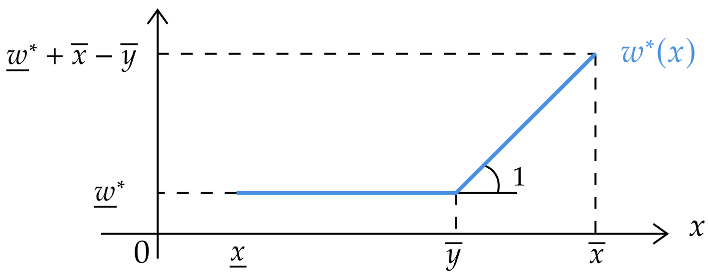
The relaxed problem can be written as an optimal control problem (P) with states (w, α, β) and control w':

$$\begin{aligned}
\min \quad & \left[\int_{\underline{x}}^{\overline{x}} w(x) f(x \mid a)\,\mathrm{d}x \right] \\
\text{subject to} \quad & w'(x) = u(x) \quad \text{for a.e. } x \\
& \alpha'(x) = k[1 - F(x \mid a)]\,u(x), \qquad \beta'(x) = k[-F_a(x \mid a)]\,u(x) \\
& w(\underline{x}) \geq 0, \quad w(\overline{x}) \text{ free} \\
& \alpha(\underline{x}) = w(\underline{x}), \quad \alpha(\overline{x}) - \psi(a) \geq U_0, \qquad \beta(\underline{x}) = 0, \quad \beta(\overline{x}) = \psi'(a) \\
& 0 \leq u(x) \leq 1 \quad \text{for all } x \in [\underline{x}, \overline{x}].
\end{aligned} \tag{P}$$

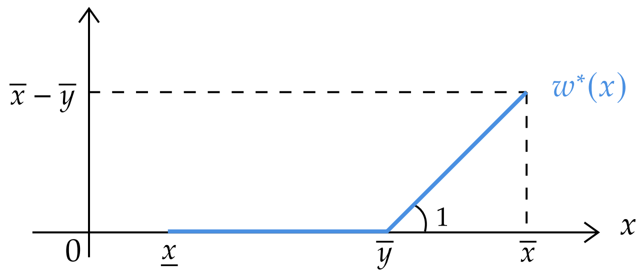
A wage schedule is called a *debt contract* if the wage is constant up to some level of outcome and grows at the same rate as outcome afterward.

Definition 2.1. Wage schedule w is a *debt contract* if there exist y such that $w'(x) = 0$ if $x \in [\underline{x}, y)$ and $w'(x) = 1$ if $x \in (y, \overline{x}]$.

[5]For action a that satisfies (2.3) with equality, the only way to induce a is to make the agent a residual claimant; that is, $w'(x) = 1$ for all x.



(i) When $\underline{w}^* > 0$



(ii) When $\underline{w}^* \leq 0$

Figure 2.2: The cost-minimizing wage schedule w^* is a debt contract.

I show that, under the strict version of the monotone likelihood ratio property, the solution to (P) for each $a \in A^*$ is a debt contract.

Assumption 2.3 (Strict MLRP)**.** For each a, $\frac{\partial}{\partial x}\left(\frac{f_a(x|a)}{f(x|a)}\right) > 0$ for all x.

Proposition 2.1. *Assume F satisfies Strict MLRP. For each $a \in A^*$, the cost-minimizing wage schedule w^* associated with the solution u^* to the relaxed problem* (P) *is debt contract. Specifically, let $\underline{w}^* \equiv U_0 + \psi(a) - \int_{\overline{y}}^{\overline{x}} k[1 - F(x \mid a)]\,\mathrm{d}x$, where $\overline{y}$ is the unique value satisfying $\int_{\overline{y}}^{\overline{x}} k[-F_a(x \mid a)]\,\mathrm{d}x = \psi'(a)$. The cost-minimizing wage schedule w^* is given by*

$$w^*(x) = \begin{cases} \max\{\underline{w}^*, 0\} & \text{if } x \in [\underline{x}, \overline{y}] \\ \max\{\underline{w}^*, 0\} + x - \overline{y} & \text{if } x \in (\overline{y}, \overline{x}]. \end{cases}$$

The solution w^* is depicted in Figure 2.2.

Given the implication of the distortion function ϕ I use, Proposition 2.1 is consistent with the literature. As I discussed at the beginning of this section, the agent behaves as if he is pessimistic, overweighting the probabilities of bad outcomes. Although he is risk-averse because of such pessimism, he can be considered "almost" risk-neutral because his net utility (2.2) is linear in wage. Indeed, Yaari (1987) discusses that risk-neutrality is characterized by ϕ being the identity function. Since the present ϕ is essentially an identity function scaled by k, it is reasonable to consider the agent almost risk-neutral In this view, Proposition 2.1 resonates with Innes (1990), who shows the optimality of debt contracts under risk-neutrality, limited liability, and monotonicity constraints.

2.3.1 Proof of Proposition 2.1

In this section, I first define a class of simple contracts to simplify the exposition of the proof. I then characterize the solution to (P). The proof consists of two steps. In the first step, using Pontryagin's maximum principle, I solve (P) for fixed base wage $w(\underline{x})$. In the second step, I solve the optimization over $w(\underline{x})$.

Debt and capped bonus contract

Recall that in (P) the principal is restricted to wage schedules that satisfy (FOA):

$$\int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)]w'(x)\,\mathrm{d}x = \psi'(a). \tag{FOA}$$

For each $a \in A^*$, by the implicit function theorem, there exists function $l_a\colon [\underline{x}, \overline{x}] \to [\underline{x}, \overline{x}]$ such that

$$\int_{y}^{l_a(y)} k[-F_a(x \mid a)]\,\mathrm{d}x = \psi'(a).$$

I define *debt and capped bonus contract* w_{DB} as follows: for any $y \in [\underline{x}, \overline{y}]$, $\underline{w}$, and a, where $\overline{y} \equiv l^{-1}(\overline{x})$

$$w_{\mathrm{DB}}(x \mid y, \underline{w}, a) = \begin{cases} \underline{w} & \text{if } x \in [\underline{x}, y] \\ \underline{w} + x - y & \text{if } x \in (y, l_a(y)] \\ \underline{w} + l_a(y) - y & \text{if } x \in [l_a(y), \overline{x}]. \end{cases}$$

That is, $w_{\mathrm{DB}}(x \mid y, \underline{w}, a)$ is the wage schedule such that (i) the base wage is $\underline{w}$; (ii) the agent receives "bonus" on $[y, l_a(y)]$; and (iii) the wage is constant outside $[y, l_a(y)]$. In

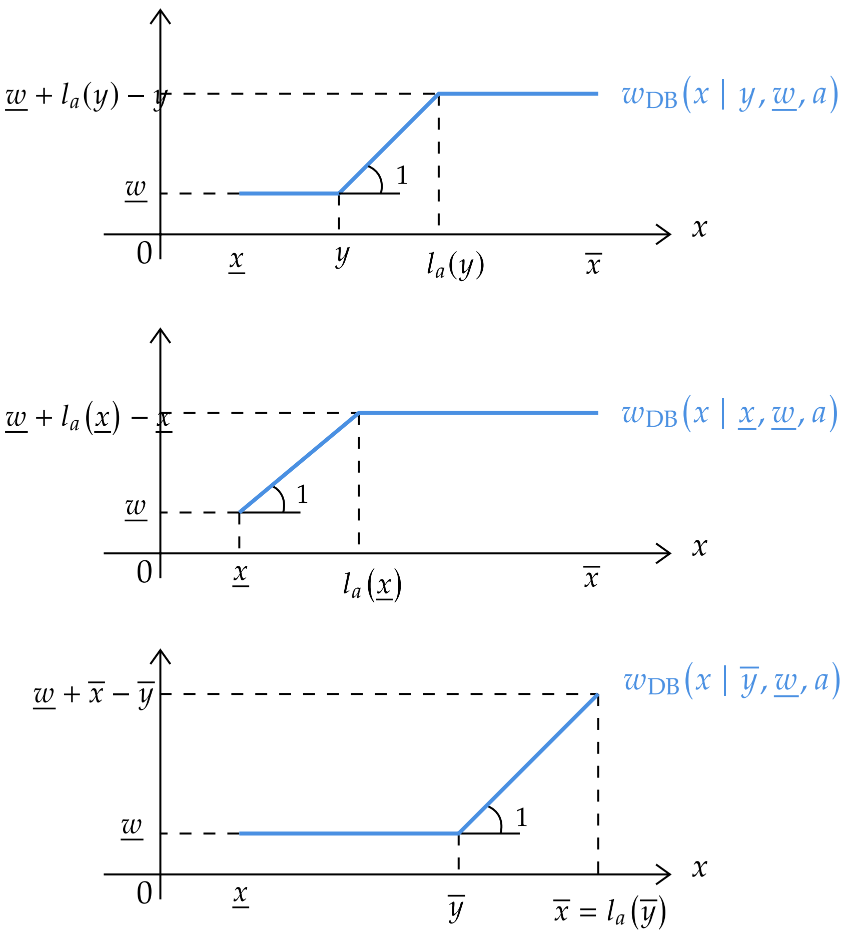


Figure 2.3: Debt and capped bonus contracts. Debt contract and capped bonus contract are special cases of this class.

particular, for any $\underline{w}$ and a, $w_{DB}(x \mid \underline{x}, \underline{w}, a)$ is a *capped bonus contract*, and $w_{DB}(x \mid \overline{y}, \underline{w}, a)$ is a *debt contract* (see Figure 2.3).

For $w_{DB}(x \mid y, \underline{w}, a)$, (PC) can be written as

$$\underline{w} \geq U_0 + \psi(a) - \int_y^{l_a(y)} k[1 - F(x \mid a)]\,\mathrm{d}x \equiv B(y).$$

Note that under Strict MLRP, the implicit function theorem also implies $B'(y) > 0$.

Proof of Proposition 2.1

First, I show that Strict MLRP implies that $\frac{f}{1-F}$ is decreasing in action, the property repeatedly used in the proof. Since $\frac{f_a}{f} = \frac{\partial}{\partial a} \ln f$, for any a and a',

$$\frac{f(x \mid a')}{f(x \mid a)} = \exp\left[\int_a^{a'} \frac{f_a(x \mid \tilde{a})}{f(x \mid \tilde{a})} \, \mathrm{d}\tilde{a}\right].$$

Hence, by Strict MLRP, for any $x < x'$ and $a < a'$,

$$\frac{f(x' \mid a')}{f(x' \mid a)} > \frac{f(x \mid a')}{f(x \mid a)} \iff f(x' \mid a')f(x \mid a) > f(x \mid a')f(x' \mid a).$$

Integrating both sides with respect to x' from x to $\overline{x}$ and rearranging,

$$\frac{f(x \mid a)}{1 - F(x \mid a)} > \frac{f(x \mid a')}{1 - F(x \mid a')} \quad \text{for all } x \in \left[\underline{x}, \overline{x}\right). \tag{2.4}$$

This proves that $\frac{f}{1-F}$ is decreasing in a.

Step 1: Solve (P) for fixed $w\left(\underline{x}\right)$

Fix $w\left(\underline{x}\right) = \underline{w} \geq 0$. Define the Hamiltonian for (P) as follows:

$$H(w, u, p_1, p_2, p_3, x) = -p_0 w f + p_1 u + p_2 k(1 - F)u + p_3 k(-F_a)u.$$

Let $u^*(x)$ be a piecewise continuous control defined on $\left[\underline{x}, \overline{x}\right]$ that solves problem (P), and let $w^*(x)$, $\alpha^*(x)$ and $\beta^*(x)$ be the associated optimal paths.[6] By Theorem 2 of Seierstad and Sydsaeter (1986, chap. 2), there exist constants p_0, p_2^*, and p_3^* and a continuous and

[6]I choose the value of $u(x)$ at a point of discontinuity x' as the left-hand limit of $u(x)$ at x'. I also assume that $u(x)$ is continuous at $\underline{x}$ and $\overline{x}$.

piecewise continuously differentiable adjoint function $p_1(x)$ such that, for all $x \in [\underline{x}, \overline{x}]$,

$$\left(p_0, p_1(x), p_2^*, p_3^*\right) \neq (0, 0, 0, 0) \tag{2.5}$$

$$p_0 = 0 \text{ or } p_0 = 1$$

$$u^*(x) \text{ maximizes } H\left(w^*(x), u, p_1(x), p_2^*, p_3^*, x\right) \text{ for all } u \text{ in } [0, 1] \tag{2.6}$$

$$\text{Except at the points of discontinuities of } u^*(x),\ p_1'(x) = p_0 f(x \mid a) \tag{2.7}$$

$$p_1(\overline{x}) = 0,\ p_2^* \geq 0 \text{ and } p_2^*\left[\alpha^*(\overline{x}) - \psi(a) - U_0\right] = 0. \tag{2.8}$$

Since the Hamiltonian is linear and hence concave in (w, u), the maximized Hamiltonian $\widehat{H}(w, p_1, p_2, p_3, x) \equiv \max_{u \in [0,1]} H(w, u, p_1, p_2, p_3, x)$ is also concave in w. Thus, by Theorem 5 of Seierstad and Sydsaeter (1986, chap. 2), the necessary conditions (2.5) to (2.8) are also sufficient for optimality if they are satisfied with $p_0 = 1$.

By (2.7) and (2.8), $p_1(x) = -p_0[1 - F(x \mid a)]$. Substituting this into the Hamiltonian,

$$H\left(w, u, p_1(x), p_2^*, p_3^*, x\right) = -p_0 w f - p_0(1 - F)u + p_2 k(1 - F)u + p_3 k(-F_a)u,$$

and hence

$$\begin{aligned} H_u^*(x) &\equiv \frac{\partial H}{\partial u}\left(w^*(x), u, p_1(x), p_2^*, p_3^*, x\right) \\ &= \left(p_2^* k - p_0\right)[1 - F(x \mid a)] + p_3^* k[-F_a(x \mid a)]. \end{aligned}$$

A necessary condition for (2.6) is

$$u^*(x) = \begin{cases} 1 & \text{if } H_u^*(x) > 0 \\ 0 & \text{if } H_u^*(x) < 0. \end{cases} \tag{2.9}$$

Suppose $p_0 = 0$. Then, $p_1(x) = 0$ for all x. Hence, by (2.5) and (2.8), at least one of $p_2^* > 0$ and $p_3^* \neq 0$ must hold.

Case 1-1: Suppose $p_2^* > 0$ and $p_3^* = 0$. Then, $H_u^*(x) = p_2^* k[1 - F(x \mid a)] > 0$ for all $x \in [\underline{x}, \overline{x})$. Hence, (2.9) implies $u^*(x) = 1$ for all x, which violates (FOA).

Case 1-2: Suppose $p_2^* = 0$ and $p_3^* \neq 0$. Since $H_u^*(x) = p_3^* k[-F_a(x \mid a)]$, either $H_u^*(x) > 0$ or $H_u^*(x) < 0$ for all $x \in (\underline{x}, \overline{x})$. Hence, (2.9) implies $u^*(x) = 1$ or $u^*(x) = 0$ for all x, which violates (FOA).

Case 1-3: Suppose $p_2^* > 0$ and $p_3^* \neq 0$. Since, $H_u^*(x) = p_2^* k[1 - F(x \mid a)] + p_3^* k[-F_a(x \mid a)]$, $H_u^*(x) > 0 \Leftrightarrow p_3^* > p_2^* \left[\frac{1-F(x|a)}{F_a(x|a)}\right] \equiv r(x)$. Note that (2.4) implies $r'(x) > 0$ and that $\lim_{x \to \underline{x}} r(x) = -\infty$. By (2.9), the unique candidate is a capped bonus contract:

$$w_{\text{DB}}(x \mid \underline{x}, \underline{w}, a) = \begin{cases} \underline{w} + x - \underline{x} & \text{if } x \in [\underline{x}, l_a(\underline{x})] \\ \underline{w} + l_a(\underline{x}) - \underline{x} & \text{if } x \in (l_a(\underline{x}), \overline{x}]. \end{cases}$$

By (2.8), (PC) must bind:

$$\underline{w} = B(\underline{x}).$$

Now suppose $p_0 = 1$. Then, $p_1(x) = -[1 - F(x \mid a)]$.

Case 2-1: Suppose $p_2^* = p_3^* = 0$. Then, $H_u^*(x) = -[1 - F(x \mid a)] < 0$ for all $x \in [\underline{x}, \overline{x})$. Hence, (2.9) implies $u^*(x) = 1$ for all x, which violates (FOA).

Case 2-2: Suppose $p_2^* > 0$ and $p_3^* = 0$. Then, $H_u^*(x) = (p_2^* k - 1)[1 - F(x \mid a)]$. If $p_2^* \neq 1/k$, then (2.9) implies either $u^*(x) = 1$ or $u^*(x) = 0$ for all x, which violates (FOA). If $p_2^* = 1/k$, then any u^* is compatible with (2.9). In this case, (PC) must bind by (2.8).

In particular, the debt and capped bonus contract $w_{\text{DB}}(x \mid y, \underline{w}, a)$ satisfy all the necessary conditions if

$$\underline{w} = B(y).$$

Case 2-3: Suppose $p_2^* = 0$ and $p_3^* \neq 0$. Since $H_u^*(x) = -[1 - F(x \mid a)] + p_3^* k[-F_a(x \mid a)]$, $H_u^*(x) > 0 \Leftrightarrow p_3^* > \frac{1-F(x|a)}{k[-F_a(x|a)]} \equiv s(x)$. Note that (2.4) implies $s'(x) < 0$ and that $\lim x \to \underline{x} = +\infty$. By (2.9), the unique candidate is a debt contract:

$$w_{\text{DB}}(x \mid \overline{y}, \underline{w}, a) = \begin{cases} \underline{w} & \text{if } x \in [\underline{x}, \overline{y}] \\ \underline{w} + x - \overline{y} & \text{if } x \in (\overline{y}, \overline{x}], \end{cases}$$

with

$$\underline{w} \geq B(\overline{y}).$$

Case 2-4: Suppose $p_2^* > 0$ and $p_3^* \neq 0$. Since, $H_u^*(x) = (p_2^* k - 1)[1 - F(x \mid a)] + p_3^* k[-F_a(x \mid a)]$, $H_u^*(x) > 0 \Leftrightarrow p_3^* > \frac{(1-p_2^* k)[1-F(x|a)]}{k[-F_a(x|a)]} \equiv t(x)$.

- If $p_2^* = 1/k$, then $t(x) = 0$ for all x. Hence, (2.9) implies $u^*(x) = 1$ or $u^*(x) = 0$ for all x, which violates (FOA).
- If $p_2^* < 1/k$, then $\lim_{x \to \underline{x}} t(x) = +\infty$ and (2.4) implies $t'(x) < 0$. By (2.9), the unique candidate is a debt contract $w_{\text{DB}}(x \mid \overline{y}, \underline{w}, a)$ with binding (PC):

$$\underline{w} = B(\overline{y}).$$

- If $p_2^* > 1/k$, then $\lim_{x \to \underline{x}} t(x) = -\infty$ and (2.4) implies $t'(x) > 0$. By (2.9), the unique candidate is a capped bonus contract $w_{\text{DB}}(x \mid \underline{x}, \underline{w}, a)$ with binding (PC):

$$\underline{w} = B(\underline{x}).$$

Recall $B'(y) > 0$. Hence, for fixed $\underline{w}$, either (i) $\underline{w} = B(\underline{x})$; (ii) $\underline{w} = B(y)$ for some $y \in (\underline{x}, \overline{y})$; or (iii) $\underline{w} \geq B(\overline{y})$. Therefore, when $w(\underline{x})$ is fixed to $\underline{w}$, an optimal path w^* associated with a solution u^* is given by

$$w^*(x) = \begin{cases} w_{\text{DB}}(x \mid \underline{x}, \underline{w}, a) & \text{if } \underline{w} = B(\underline{x}) \\ w_{\text{DB}}(x \mid y, \underline{w}, a) & \text{if } \underline{w} = B(y) \\ w_{\text{DB}}(x \mid \overline{y}, \underline{w}, a) & \text{if } \underline{w} \geq B(\overline{y}). \end{cases} \tag{2.10}$$

Step 2: Optimization over $w(\underline{x})$

Given (2.10), I now solve the optimization over $\underline{w}$. Note that among the debt contracts $w_{\text{DB}}(x \mid \overline{y}, \underline{w}, a)$, the one with $\underline{w} = B(\overline{y})$ minimizes the expected payment to the agent. Hence, it is without loss to focus on $w_{\text{DB}}(x \mid y, B(y), a)$, and I consider the optimal choice of y for the principal. The principal's expected payment under this contract is

$$\begin{aligned}
&\int_{\underline{x}}^{y} B(y)f(x \mid a)\,\mathrm{d}x + \int_{y}^{l_a(y)} \big[B(y) + x - y\big] f(x \mid a)\,\mathrm{d}x + \int_{l_a(y)}^{\overline{x}} \big[B(y) + l_a(y) - y\big] f(x \mid a)\,\mathrm{d}x \\
&= U_0 + \psi(a) - \int_{y}^{l_a(y)} k[1 - F(x \mid a)]\,\mathrm{d}x + \int_{y}^{l_a(y)} x f(x \mid a)\,\mathrm{d}x \\
&\qquad - y\big[1 - F(y \mid a)\big] + l_a(y)\big[1 - F\big(l_a(y) \mid a\big)\big] \\
&= U_0 + \psi(a) + (1-k)\int_{y}^{l_a(y)} [1 - F(x \mid a)]\,\mathrm{d}x,
\end{aligned}$$

where the second inequality follows from applying integration by parts to the fourth term. The implicit function theorem implies that the third term in the last line is decreasing in y. Thus, the principal always chooses a debt contract ($y = \overline{y}$). Letting $\underline{w}^* = B(\overline{y})$, the cost-minimizing wage schedule w^* is given by

$$w^*(x) = w_{\mathrm{DB}}\big(x \mid \overline{y}, \max\{\underline{w}^*, 0\}, a\big) = \begin{cases} \max\{\underline{w}^*, 0\} & \text{if } x \in [\underline{x}, \overline{y}] \\ \max\{\underline{w}^*, 0\} + x - \overline{y} & \text{if } x \in (\overline{y}, \overline{x}]. \end{cases}$$

2.4 The validity of the first-order approach

I now turn to the validity of the FOA. Leveraging the technique in Chade and Swinkels (2020), I derive a sufficient condition under which the agent's net utility is strictly quasiconcave in action, which implies that the condition (FOA) characterizes the optimal choice of action. As in Chade and Swinkels (2020), the concavities (in proportionate terms) of outcome distribution F and disutility ψ are the keys to validating the FOA.

Proposition 2.2. *Under Strict MLRP, the first-order approach is valid if*

$$\frac{F_{aa}(x \mid a)}{F_a(x \mid a)} < \frac{\psi''(a)}{\psi'(a)} \qquad \textit{for all } x \in (\overline{y}, \overline{x}). \tag{2.11}$$

Proof. The proof follows almost the same steps as Chade and Swinkels (2020, Proposition

5). Recall that the agent's net utility from wage schedule w and action $\tilde{a}$ is given by

$$U(w,\tilde{a}) = w\left(\underline{x}\right) + \int_{\underline{x}}^{\overline{x}} k[1 - F(x \mid \tilde{a})]\, w'(x)\, \mathrm{d}x - \psi(\tilde{a}).$$

The first- and second-order derivatives with respect to action are

$$\begin{aligned} U_a(w,\tilde{a}) &= \int_{\underline{x}}^{\overline{x}} k[-F_a(x \mid a)]w'(x)\, \mathrm{d}x - \psi'(a) \\ U_{aa}(w,\tilde{a}) &= \int_{\underline{x}}^{\overline{x}} k[-F_{aa}(x \mid a)]w'(x)\, \mathrm{d}x - \psi''(a), \end{aligned}$$

If $U_a(w,a) = 0$,

$$\psi''(a) = \frac{\psi''(a)}{\psi'(a)}\, \psi'(a) = \int_{\underline{x}}^{\overline{x}} \frac{\psi''(a)}{\psi'(a)} k[-F_a(x \mid a)]w'(x)\, \mathrm{d}x.$$

Using this equation, $U_{aa}(w,a)$ can be written as

$$U_{aa}(w,a) = -k \int_{\underline{x}}^{\overline{x}} \left[F_{aa}(x \mid a) - \frac{\psi''(a)}{\psi'(a)} F_a(x \mid a) \right] w'(x)\, \mathrm{d}x.$$

Note that $U_{aa}(w,a) < 0$ suffices for $U(w,\cdot)$ to be strictly quasiconcave. For w^*, this condition is reduced to

$$\int_{\overline{y}}^{\overline{x}} \left[F_{aa}(x \mid a) - \frac{\psi''(a)}{\psi'(a)} F_a(x \mid a) \right] \mathrm{d}x > 0,$$

which holds if

$$\frac{F_{aa}(x \mid a)}{F_a(x \mid a)} < \frac{\psi''(a)}{\psi'(a)} \qquad \text{for all } x \in \left(\overline{y}, \overline{x}\right).$$

□

The condition (2.11) requires that for all $x \in \left(\overline{y}, \overline{x}\right)$, the outcome distribution F is (in proportionate terms) less convex in action than the disutility ψ. In particular, (2.11) is satisfied if both F and ψ are convex in action, as in Rogerson (1985).

2.5 Discussion

In this section, I discuss some alternative modeling assumptions.

2.5.1 Distortion function ϕ

While Yaari (1987) characterizes dual risk-aversion by the convexity of ϕ, I have so far assumed a specific convex and *piecewise-linear* ϕ that has a natural interpretation. Another simplest possible ϕ is a piecewise-linear function that stays at 0 until q and extends from $(q, 0)$ to $(1, 1)$:

$$\phi(p) = \begin{cases} 0 & \text{if } 0 \leq p \leq q \\ \frac{p-q}{1-q} & \text{if } q < p \leq 1 \end{cases}$$

for some $q \in (0, 1)$. See Figure 2.4. This ϕ also has a natural interpretation. It is easy to see that the agent's gross utility can be written as

$$w(\underline{x}) + \int_{\underline{x}}^{x(a)} \frac{1 - F(x \mid a) - q}{1 - q} w'(x) \, \mathrm{d}x,$$

where $x(a) \equiv F^{-1}(1 - q \mid a)$. That is, the agent behaves as if he is so pessimistic that he ignores the possibilities of good outcomes $x > x(a)$. Note that a debt contract can never be optimal in this case; since the agent virtually ignores outcomes $[x(a), \overline{x}]$, it is optimal for the principal to set a constant wage in this interval. Indeed, by following similar steps as in Section 2.3, I can show that the cost-minimizing wage schedule features a constant wage for high (and low) outcomes, and that the wage grows at the same rate with outcome elsewhere.

This observation can be further generalized to a general convex ϕ. Specifically, I can show that the cost minimizing wage schedule must have bang-bang property; that is, the slope $w'(x)$ is either 0 or 1 for almost every x. Bang-bang property results from the fact that the Hamiltonian is linear in control, whatever ϕ I use, and hence is a robust property of the solution.

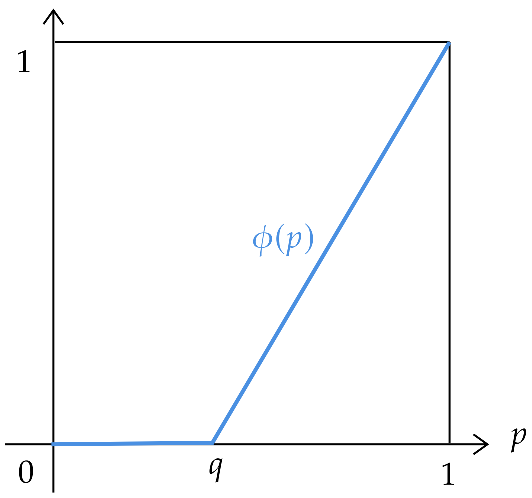


Figure 2.4: Another piecewise-linear distortion ϕ.

2.5.2 Bounded slope

Following Poblete and Spulber (2012), I have assumed Monotonicity and Bounded Slope of wage schedules (Assumptions 2.1 and 2.2). On one hand, they have some rationales such as deterring sabotage by either the principal or the agent. On the other hand, they jointly imply Lipschitz continuity of wage schedules, thereby excluding any discontinuity. In particular, wage schedules with upward jumps are not admissible.

The principal's problem is well defined even without the Bounded Slope assumption. A simple calculation can show that a two-part wage schedule, which pays a flat base-wage until the threshold and pays a flat bonus-wage afterward, can induce the same action at a smaller cost than the cost-minimizing schedule characterized in Section 2.3. Now that upward jumps are admissible, by letting the wage jump up at a certain threshold, the principal can give the agent enough incentive without creating a "bonus region" as in the case of a debt contract. This helps reduce the expected payment to the agent, thereby achieving a smaller agency cost.

2.6 Conclusion

This chapter has explored the implications of dual risk-aversion in a standard moral hazard setting with monotonicity and limited liability constraints. Under a specific form of risk-aversion where the agent overweights the probability of bad outcomes, the cost-minimizing wage schedule takes a particularly simple form, a debt contract. Given that the agent is considered almost risk-neutral, this result resonates with Innes (1990), who shows the optimality of debt contracts under similar constraints. Leveraging the technique developed in Chade and Swinkels (2020), I have also derived a sufficient condition for validating the FOA. As is consistent with Chade and Swinkels (2020), the condition involves the concavities (in proportionate terms) of outcome distribution and disutility of action and is implied by Rogerson (1985)'s Convexity of Distribution Fucntion Condtion and convex disutility.

There are some extensions worth exploring. First, it is of theoretical interest to fully characterize the shape of the cost-minimizing wage schedule under a general distortion function. Although the bang-bang property can be easily shown, it is not yet clear how dual risk-aversion interacts with the principal's incentive and shapes the cost-minimizing wage schedule. Second, I have assumed the monotonicity of wage schedules throughout the chapter. It has some rationale, such as preventing the agent from sabotaging the task. However, the monotonicity of the solution is indeed one of the results in the seminal work of Holmström (1979) and the follow-up papers. It is thus worthwhile to explore whether the monotonicity can be obtained through the principal's optimization problem.

Chapter 3

Robust Predictions and Hard Information in the Market for Lemons

3.1 Introduction

Informationally robust predictions aim to identify the set of outcomes that can arise in (Bayesian) equilibrium under *some* admissible information structure. The literature on such predictions has focused exclusively on *soft* information; that is, information that cannot be credibly verified or disclosed to others. Roesler and Szentes (2017) and Bergemann, Brooks and Morris (2015) study stylized trading environments between a monopoly seller and buyer(s). Roesler and Szentes (2017) show that the optimal information structure for the buyer is not (necessarily) the one in which he knows everything; it may be better for him not to know his precise valuation. Bergemann et al. (2015) characterize the entire set of possible equilibrium payoffs, assuming that buyers know their precise valuations while the seller may have imprecise information about them. Kartik and Zhong (2024) further extend this approach to settings with interdependent values, maintaining the assumption that information is soft.

In practice, however, information is often *hard*; that is, it is accompanied by *evidence* and can be credibly disclosed to others. For example, a car manufacturer obtains a third-party certification for the quality of its car. As is well understood in the literature on hard information, such information may be elicitable without satisfying the same sort of incentive compatibility as soft information.[1] Therefore, the set of possible trading outcomes could be different if we allow for soft *and* hard information.

In this chapter, we show in a stylized setup of adverse selection (Lemon) market à la Akerlof (1970) that such hard information is indeed critical in informationally robust prediction. The buyer is assumed to have no information (as the seller in Roesler and Szentes, 2017). The seller may have any (possibly noisy) information about the quality. First, we consider an arbitrary "soft" information structure; that is, the seller may obtain an arbitrary signal about the quality of the good, but no evidence is attached to it. We show that, under certain conditions, the unique equilibrium outcome is no-trade regardless of how rich the seller's (soft) information structure is. Next, we demonstrate by example that, with hard information, non-trivial trade outcomes could be supported. Moreover, the difference is economically significant because the players' payoffs and total surplus are strictly improving over the no-trade outcome. These results highlight the importance of allowing for both soft *and* hard information when studying informationally robust predictions.

Our model differs from Ali, Kleiner and Zhang (2024) and Dasgupta, Krasikov and Lamba (2022), two recent papers that study hard information design in standard monopoly pricing settings, by assuming an initially uninformed sender (seller), interdependent values, and no commitment on the part of the receiver (buyer). In our model, both players are initially uninformed, and the seller learns her type through either soft or hard information. By contrast, Ali et al. (2024) consider a setting in which the sender knows his type and strategically discloses the hard information he acquires to influence the receiver's behavior. Dasgupta et al. (2022) is closer to our model in that the sender is initially uninformed and learns his type through hard information. They show that hard information can expand the set of implementable outcomes. However, they assume private values and the receiver's commitment to a pricing scheme, whereas we assume interdependent values and no commitment on the part of the receiver.

[1] See Green and Laffont (1986) and Bull and Watson (2007).

In the adverse selection literature, the importance of hard information has long been recognized (e.g., see Viscusi, 1978), though much of the work assumes some fixed information structure. Our contribution is to highlight its importance even for informationally robust predictions, and in the environment beyond those previously studied in the literature. We hope the present work can be used as a small but first step toward a fuller understanding of this subject.

3.2 Model

There is a seller (she) who has a good to sell to a buyer (he). There are n possible quality levels (types) that the good can take, denoted by $q \in \{1, 2, \ldots, n\}$. The common prior over types is given by the CDF F. If the good is of type q, the seller's valuation is $r_q \in \mathbb{R}_+$ and the buyer's valuation is $v_q \in \mathbb{R}_+$. We order the types such that $r_1 < r_2 < \cdots < r_n$ and $v_1 < v_2 < \cdots < v_n$. We assume that the valuations are linearly related and that trade is ex post efficient for the highest type but inefficient for the lowest one; specifically, $v_1 < r_1$ and $r_n < v_n$. The seller observes a signal s about q. The buyer then makes a take-it-or-leave-it price offer of p to the seller.[2] Trade occurs if and only if the seller's expected valuation conditional on the realized signal weakly exceeds p.[3] If type q good is traded at price p, the seller gets $p - r_q$ and the buyer gets $v_q - p$; if no trade occurs, both get 0.

Given the linear relationship between valuations, we can, without loss of generality, assume that the seller observes an *unbiased signal*; that is, $\mathbb{E}[r \mid s] = s$. Under this assumption, the buyer's posterior mean is a linear function of s:

$$v(s) \equiv \mathbb{E}[v \mid s] = \frac{v_n - v_1}{r_n - r_1} s - \frac{r_1 v_n - v_1 r_n}{r_n - r_1}.$$

[2]The assumption that the (uninformed) buyer is an offerer makes the problem simpler, by avoiding the informed-principal problem. Though not innocuous, it seems a reasonable assumption in some contexts, such as when the (uninformed) buyer is a potential acquirer of a firm who often has a stronger bargaining power than the (informed) seller.

[3]We take this seller's acceptance criterion for granted to simplify the analysis, but it is not crucial for the results.

Let $r_F \equiv \int_{r_1}^{r_n} r \, \mathrm{d}F(r)$ denote the seller's expected valuation under the prior. We assume that no trade occurs under the prior:

$$v(r_F) < r_F \iff r_F < \frac{r_1 v_n - v_1 r_n}{v_n - r_n + r_1 - v_1} \equiv \underline{p}.$$

As in Roesler and Szentes (2017), a CDF G is the distribution of some unbiased signal if and only if F is a mean-preserving spread of G. Hence, a signal distribution G is feasible if it satisfies

$$\int_{r_1}^{t} F(r) \, \mathrm{d}r \geq \int_{r_1}^{t} G(s) \, \mathrm{d}s \text{ for all } t \in [r_1, r_n] \quad \text{and} \quad \int_{r_1}^{r_n} s \, \mathrm{d}G(s) = r_F.$$

3.3 Only soft information

Suppose that the seller can acquire only soft information. Proposition 3.1 demonstrates that such information acquisition does not mitigate the equilibrium inefficiency.

Proposition 3.1. *There exists no signal distribution under which trade occurs in equilibrium.*

Proof. Consider an arbitrary feasible signal distribution G. If the buyer offers price p, the seller accepts it if $s \leq p$, which yields the buyer's expected payoff of

$$\int_{r_1}^{p} \left[v(s) - p \right] \mathrm{d}G(s).$$

Since $v(\cdot)$ is increasing and $v(p) - p \leq 0$ for all $p \leq \underline{p}$, the buyer's expected payoff is negative if he offers $p \in \left[r_l, \underline{p} \right]$. If he offers $p \in \left(\underline{p}, r_h \right]$ satisfying $G(p) > 0$, his expected payoff is

$$\begin{aligned}
&\int_{r_1}^{p} \left[\frac{v_n - v_1}{r_n - r_1} s - \frac{r_1 v_n - v_1 r_n}{r_n - r_1} - p \right] \mathrm{d}G(s) \\
&= \frac{v_n - v_1}{r_n - r_1} \int_{r_1}^{p} s \, \mathrm{d}G(s) - \left[\frac{r_1 v_n - v_1 r_n}{r_n - r_1} + p \right] \int_{r_1}^{p} \mathrm{d}G(s) \\
&= G(p) \left\{ \frac{v_n - v_1}{r_n - r_1} \mathbb{E}\left[s \mid r_1 \leq s \leq p \right] - \frac{r_1 v_n - v_1 r_n}{r_n - r_1} - p \right\}.
\end{aligned}$$

This is negative if $r_F < \frac{r_1 v_n - v_1 r_n}{v_n - v_1} + \frac{r_n - r_1}{v_n - v_1} p$. Hence, it suffices to have

$$r_F \leq \frac{r_1 v_n - v_1 r_n}{v_n - v_1} + \frac{r_n - r_1}{v_n - v_1} \underline{p} = \underline{p}.$$

Since $r_F < \underline{p}$ by assumption, the unique equilibrium outcome is no-trade, whatever signal distribution we consider. □

3.4 Soft and hard information

Suppose now that the seller can obtain and disclose hard information about her type. Specifically, upon observing a signal s, the seller obtains a set of evidence $M(s)$. She then discloses some $m \in M(s)$ to the buyer, after which the buyer makes an offer. We continue to assume that the seller accepts the offer if and only if her expected valuation conditional on the realized signal (and the set of evidence) weakly exceeds the offer. We demonstrate by example that, in equilibrium, trade can occur and the seller can obtain a positive expected payoff.

Example 1. Assume $n = 2$ and $0 \leq v_1 < r_1 < r_2 < v_2$. We maintain the assumption that no trade occurs under the prior: $v_F \equiv v(r_F) < r_F$. Consider a signal distribution G such that, for some $g \in \left(\frac{(r_F - r_1)(r_2 - v_F)}{(v_2 - r_2)(r_F - r_1) + (r_2 - r_1)(r_2 - v_F)}, \frac{r_F - r_1}{r_2 - r_1} \right)$, the signal s is distributed as

$$s = \begin{cases} r_1 & \text{with probability } \frac{r_2 - r_F}{r_F - r_1} g \\ r_F & \text{with probability } 1 - \frac{r_2 - r_1}{r_F - r_1} g \\ r_2 & \text{with probability } g. \end{cases}$$

It is straightforward to verify that the common prior F is a mean-preserving spread of G.[4] Suppose that the set of evidence is given by $M(r_1) = \{x\}$ and $M(r_F) = M(r_2) = \{x, y\}$; that is, if the seller observes $s = r_1$, she can only disclose $m = x$, whereas if she observes $s = r_F$ or $s = r_2$, she disclose either $m = x$ or $m = y$ This evidence structure can be

[4]Indeed, $\mathbb{E}_G[s] = r_F$, and the area between F and G on the interval $[r_1, r_F]$ is the equals that on $[r_F, r_2]$, which establishes the required inequality.

interpreted as a certification system; only when the seller receives the signal r_F or r_2 does she obtain the certification y and can disclose it to the buyer. We show that the following strategy profile, together with the belief system derived by the profile and G through Bayes' rule, constitutes a weak perfect Bayesian equilibrium:

- Seller's strategy:
    1. Discloses $m = x$ if she observes $s = r_1$;
    2. Discloses $m = y$ if she observes $s = r_F$ or $s = r_2$.
- Buyer's strategy:
    1. Offers some $p \in [0, r_1)$ if he receives $m = x$;
    2. Offers $p = r_2$ if he receives $m = y$.

Given the buyer's strategy, the sequential rationality of the seller's strategy is obvious. The first part of the buyer's strategy is also optimal since, upon receiving $m = x$, he believes $s = r_1$ with probability 1, and no gain from trade exists in this case. It remains to verify the optimality of the second part of the buyer's strategy. Suppose that the buyer receives $m = y$. If he offers $p \in [r_F, r_2)$, the seller accepts only when $s = r_F$, yielding a negative expected payoff to the buyer because $v_F < r_F$ by assumption. If he offers $p \geq r_2$, the seller accepts regardless of whether $s = r_F$ or $s = r_2$. Hence, the buyer's optimal offer is $p = r_2$, which yields an expected payoff of

$$\frac{1}{1 - \frac{r_2 - r_F}{r_F - r_1} g} \left[g(v_2 - r_2) + \left(1 - \frac{r_2 - r_1}{r_F - r_1} g\right)(v_F - r_2) \right],$$

which is positive if

$$g(v_2 - r_2) + \left(1 - \frac{r_2 - r_1}{r_F - r_1} g\right)(v_F - r_2) > 0 \iff g > \frac{(r_F - r_1)(r_2 - v_F)}{(v_2 - r_2)(r_F - r_1) + (r_2 - r_1)(r_2 - v_F)}.$$

Therefore, offering $p = r_2$ is optimal for the buyer. Note that trade occurs with positive probability and that the seller obtains a positive expected payoff. Thus, the availability of hard information mitigates the equilibrium inefficiency.

Bibliography

Akerlof, George A. (1970) "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, Vol. 84, No. 3, pp. 488–500.

Ali, S. Nageeb, Andreas Kleiner, and Kun Zhang (2024) "From Design to Disclosure," mimeo.

Bergemann, Dirk, Benjamin Brooks, and Stephen Morris (2015) "The Limits of Price Discrimination," *American Economic Review*, Vol. 105, No. 3, pp. 921–57.

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.

Bull, Jesse and Joel Watson (2007) "Hard evidence and mechanism design," *Games and Economic Behavior*, Vol. 58, No. 1, pp. 75–93.

Carroll, Gabriel (2015) "Robustness and Linear Contracts," *American Economic Review*, Vol. 105, No. 2, pp. 536–63.

Chade, Hector and Jeroen Swinkels (2020) "The no-upward-crossing condition, comparative statics, and the moral-hazard problem," *Theoretical Economics*, Vol. 15, No. 2, pp. 445–476.

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.

Dasgupta, Sulagna, Ilia Krasikov, and Rohit Lamba (2022) "Hard information design," mimeo.

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 (2012) "*Simple* Mediation in a Cheap-Talk Game," April, Working Paper, University of Birmingham.

Gershkov, Alex, Benny Moldovanu, Philipp Strack, and Mengxi Zhang (2021) "Optimal Auctions: Non-expected Utility and Constant Risk Aversion," *The Review of Economic Studies*, Vol. 89, No. 5, pp. 2630–2662.

——— (2023) "Optimal Insurance: Dual Utility, Random Losses, and Adverse Selection," *American Economic Review*, Vol. 113, No. 10, pp. 2581—2614.

——— (2025) "Optimal Security Design for Risk-Averse Investors," *American Economic Review*, Vol. 115, No. 6, pp. 2050—92.

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.

Green, Jerry R. and Jean-Jacques Laffont (1986) "Partially Verifiable Information and Mechanism Design," *The Review of Economic Studies*, Vol. 53, No. 3, pp. 447–456.

Holmström, Bengt (1979) "Moral Hazard and Observability," *The Bell Journal of Economics*, Vol. 10, No. 1, pp. 74–91.

Holmström, Bengt and Paul Milgrom (1987) "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica*, Vol. 55, No. 2, pp. 303–328.

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

Innes, Robert D (1990) "Limited liability and incentive contracting with ex-ante action choices," *Journal of Economic Theory*, Vol. 52, No. 1, pp. 45–67.

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.

Jewitt, Ian (1988) "Justifying the First-Order Approach to Principal-Agent Problems," *Econometrica*, Vol. 56, No. 5, pp. 1177–1190.

Kamin, Debra (2024) "Buying or Selling a Home? The Rules Have Changed.," *The New York Times*.

Kartik, Navin and Weijie Zhong (2024) "Lemonade from Lemons: Information Design and Adverse Selection," mimeo.

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 (2021) "A Mediator Approach to Mechanism Design with Limited Commitment," November, 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.

Poblete, Joaquín and Daniel Spulber (2012) "The form of incentive contracts: agency with moral hazard, risk neutrality, and limited liability," *The RAND Journal of Economics*, Vol. 43, No. 2, pp. 215–234.

Roesler, Anne-Katrin and Balázs Szentes (2017) "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, Vol. 107, No. 7, pp. 2072–80.

Rogerson, William P. (1985) "The First-Order Approach to Principal-Agent Problems," *Econometrica*, Vol. 53, No. 6, pp. 1357–1367.

Seierstad, Atle and Knut Sydsaeter (1986) *Optimal control theory with economic applications*: Elsevier North-Holland, Inc.

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

Viscusi, W. Kip (1978) "A Note on "Lemons" Markets with Quality Certification," *The Bell Journal of Economics*, Vol. 9, No. 1, pp. 277–279.

Yaari, Menahem E. (1987) "The Dual Theory of Choice under Risk," *Econometrica*, Vol. 55, No. 1, pp. 95–115.