HOW INFORMED DO YOU WANT YOUR PRINCIPAL TO BE?

RAHUL DEB[◊], MALLESH PAI[†], AND ANNE-KATRIN ROESLER[♚]

NOVEMBER 15, 2023

ABSTRACT: We study a bilateral trade setting with interdependent values and two-sided private information. A buyer's value for a good depends both on his privately known type and the good's quality that he does not observe. The cost of the seller also depends on both the buyer's type and the quality; she learns about the latter via a private signal whose realization is the seller's private information. How much (if any) private information would the buyer want the seller to have? We answer this question by characterizing the buyer-optimal outcome: this is the information structure and the corresponding seller-optimal equilibrium of the informed principal game that yield the highest consumer surplus. We show that, and characterize precisely when, private information for the seller leads to higher consumer surplus relative to an uninformed seller. Under these conditions, the information structure in the buyer-optimal outcome is typically noisy.

1. INTRODUCTION

The bilateral trade of many different goods involves buyers and sellers who have interdependent values and private information. For instance, consider the sale of a house. A buyer privately knows the importance he places on different features of the house and the neighborhood. Conversely, the seller of the house may have private information about the quality of the house. Specifically, she may be aware of repairs that will soon become necessary, the noisiness of the neighbors etc. The buyer's value for the house clearly depends not just on his own but also the seller's private information. Conversely, the seller's value of the house depends on its quality (but typically not the type of the buyer he sells to). The seller's terms may depend on, and therefore signal, their private information; consequently, this is an *informed principal setting* in the sense of (Myerson, 1983). The purpose of this paper is to answer the question posed in the title: How informed do you want the principal to be? In the context of this example, the question amounts to whether the buyer is better off if the seller knows what is wrong (or right) with the house. If seller private information can indeed be beneficial to the buyer, how much private information leads to the highest consumer surplus?

[◊]DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO, RAHUL.DEB@UTORONTO.CA
[†]DEPARTMENT OF ECONOMICS, RICE UNIVERSITY, MALLESH.PAI@RICE.EDU
[♚]DEPARTMENT OF ECONOMICS, UNIVERSITY OF TORONTO, AK.ROESLER@UTORONTO.CA

We are grateful for comments from Nageeb Ali, Gabriel Carroll, Hector Chade, Jan Knoepfle, Roger Myerson, Vasiliki Skreta and numerous conference, seminar participants. Deb and Roesler thank the Social Sciences and Humanities Research Council for their continued and generous financial support. The most recent version of this paper is available here.

As an alternate example, consider a health insurance provider selling a contract to a buyer seeking coverage. The buyer's private information about his health, habits and lifestyle is relevant to both their value for having health insurance, and the seller's cost of providing coverage. Conversely, the seller may have private information about the buyer's health risks that the buyer himself may not be aware of. For instance, the seller may have access to hospital and population health data, and can use that along with the buyer's demographic characteristics to determine their health risks. Therefore, both the value to the buyer and the cost of the insurance contract to the seller depend jointly on the private information of both parties. Once again, the seller's private information makes this an informed principal setting.

Our focus on consumer surplus (as suggested by the title) stems from the rising concern about the impact of big data on consumers through the sellers' ability to use information about buyers.[1] These concerns have raised regulatory interest in limiting the information that sellers can acquire and use in pricing, either directly or by mandating the use of technologies such as differential privacy (Dwork and Roth, 2014). In the health insurance example above, the seller may be disallowed from using certain buyer attributes in designing an insurance offer, or even potentially disallowed from making distinct offers to different buyers. Clearly such regulations affect the terms of trade offered by the seller. We ask whether, and when, such regulations are well-founded from the perspective of buyer welfare.

We study a (in our opinion canonical) model in which a seller of a good of binary quality (high or low) determines the terms of trade with a buyer whose privately known type is also binary (and also high or low). The buyer value and seller cost both depend on the buyer type and the good's quality. The seller learns about the quality of the good by privately observing a signal realization from a given information structure. Upon observing the signal realization, the seller offers a mechanism to the buyer (that takes as inputs messages from both the buyer and seller). If buyer accepts the mechanism, he and the seller make simultaneous reports to the mechanism which in turn determines the probability of trade and the transfer. This informed principal game (like most signaling games) has multiple equilibria. We focus on the *seller-optimal equilibrium* that yields the highest expected profit for the seller (evaluated ex ante before the seller receives her signal realization).

We term an information structure (from which the seller learns the quality) and a seller-optimal equilibrium to be an outcome. Our aim is to derive the properties of the *buyer-optimal outcome*: this is an outcome that yields the highest consumer surplus. A first intuition might be that the buyer-optimal outcome always features an (uninformative) information structure which provides the seller with no private information. After all, one might expect that seller-private information can only hurt the buyer since this allows the seller to potentially use this information to her advantage. We show that this is not the case. We characterize the conditions under which the information structure in the buyer-optimal outcome is informative and, moreover, we show that typically, this information is noisy.

[1]For a summary of the issues and existing regulatory frameworks, we refer the reader to the US Council of Economic Advisers' 2015 report on "Big Data and Differential Pricing," `https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf`.

At a high-level, our central insight is simple: endowing the seller with private information changes the pattern of trade (i.e., the probability of trade between different buyer types and seller qualities). In particular, the seller optimal equilibrium may increase the probability of trade with low type buyers, which will be accompanied by higher consumer surplus in the form of information rents. Specifically, suppose that, absent any private information, it is not profitable for the seller to trade with the low-value buyer. Since the buyer's type is binary, this implies that the profit-maximizing seller will offer a price equal to the expected value of the high-value buyer and the consumer surplus will be zero. If private information makes it profitable for the seller to trade with the low-value buyer for certain signal realizations, the buyer will receive a strictly positive consumer surplus since the high-value buyer will earn information rents with positive probability. Our main result characterizes precisely when this occurs and the information structure that maximizes the probability of trade with the low-value buyer. As we show, this will lead to additional gains from trade while leaving the seller with the minimal surplus share in this case, thereby maximizing consumer surplus (via information rents).

Our main result therefore provides a clear economic message: Some restrictions should typically be placed on sellers but a well-intentioned, intuitive regulation preventing sellers from using any private information whatsoever might make buyers worse off and be self-defeating.

1.1. *Related Literature*

This paper is related to three distinct strands of the literature. The first is the literature on mechanism design with an informed principal following (Myerson, 1983). Like us, some readers might find it surprising that the seller-optimal equilibrium in a seemingly canonical two-type interdependent value setting such as ours has not previously been derived. The informed principal problem is hard to solve at a high degree of generality and consequently the literature has focused on the case of private values (Maskin and Tirole, 1990; Mylovanov and Tröger, 2014), common values (Maskin and Tirole, 1992) or settings where the agent's value depends on the principal's private information but the principal's private information does not influence her own cost (Koessler and Skreta, 2016). Most recently, Nishimura (2022) studies a general interdependent value setting but restricts attention to what are known as "Rothschild-Stiglitz-Wilson" allocations. We discuss the relationship to this paper in slightly greater detail after we present our first result but it is worth mentioning here that this restriction is substantial and that seller-optimal equilibria do not typically have Rothschild-Stiglitz-Wilson allocations.

The second related strand is the burgeoning literature on information design: recent surveys are Bergemann and Morris (2019) and Kamenica (2019). Within this literature we are closest to two papers. The first is Roesler and Szentes (2017) who introduce the problem of deriving the buyer-optimal outcome in a standard monopoly setting (without seller private information). The second is Kartik and Zhong (2023). They first study a common value environment (Ìa Akerlof, 1970) with the aim of characterizing all possible combinations of the producer and consumer surplus (in the spirit of Bergemann, Brooks, and Morris, 2015) that can arise when both the buyer and the seller learn about the common value parameter (recall, by contrast, that our buyer is perfectly informed about his type). In this setting, they show that all possible consumer-producer surplus pairs can arise by considering different equilibria for the trivial information structure where both players learn nothing. They argue that this insight generalizes to a multidimensional environment like ours provided that the seller is restricted to posting prices. Both differences (buyer learning and the restriction of seller mechanisms) are substantive because as we have mentioned above, for certain parameter values, our buyer-optimal outcome cannot feature an uninformed seller and, moreover, in this outcome, the seller does not offer a posted price.

The final strand of the related literature lies at the intersection of economics and computer science. This work aims to understand the value of (various forms of) buyer privacy in strategic settings. A majority of this literature considers the case of repeated sales where privacy corresponds to whether the seller remembers the buyer in subsequent interaction. The overwhelming message is the surprising finding that buyer privacy may hurt the buyer: for instance, see Conitzer, Taylor, and Wagman (2012) or Cummings, Ligett, Pai, and Roth (2016) for specific applied settings and Calzolari and Pavan (2006) for a general result in a sequential contracting setting. Recovering a positive value of privacy in such settings generally requires buyers to be behavioral or not fully strategic; a classic reference is Taylor (2004). The closest paper is the contemporaneous work of Brunnermeier, Lamba, and Segura-Rodriguez (2021): they study the value of buyer privacy in insurance settings and, like us, the buyer's value depends on both parties' private information. Unlike our standard Bayesian framework however, their (behavioral) buyer does not fully update (about the seller's private information) from the seller's offer, and therefore can be "exploited."

2. Model

We study a canonical (binary types) interdependent value setting in which a seller (she) has a single good to sell to a buyer (he) and both have private information. In this section, we present a benchmark model to simplify the presentation and we discuss extensions in Section 5.

2.1. *Values*

The buyer has a privately known *type* $\theta \in \Theta := \{\theta_h, \theta_\ell\}$ that captures the extent to which the buyer values the features of the good; θ_h, θ_ℓ are drawn with commonly known probabilities $f_h \in (0,1)$, $f_\ell = 1 - f_h$, respectively.

The good has a *quality* $q \in Q := \{q_h, q_\ell\}$; q_h, q_ℓ are drawn (independently from the buyer's type) with commonly known probabilities $p_h \in (0,1)$, $p_\ell = 1 - p_h$ respectively.[2] This quality is unknown to the buyer. As we describe below, the seller privately observes an informative signal about the quality.

The buyer's utility depends on both the buyer's type and the good's quality. For now, we assume that the seller's cost depends only on the quality; in Section 5 we present the extension in which the seller's cost can also depend on the buyer type. We use $U_{bs} > 0$ and $C_s > 0$ to denote the *buyer's value* and *seller's cost* respectively when the buyer's type is θ_b and the quality of the good is q_s. Since our environment is binary, the values can be summarized in the following convenient matrix form (Table 1).

We make the following assumptions on the values.

(1) *Buyer value monotonicity*: $U_{hs} > U_{\ell s}$ for $s \in \{\ell, h\}$ and $U_{bh} > u_{b\ell}$ for $b \in \{\ell, h\}$.

[2]While we refer to q the quality of the good, it can also capture any other attribute of the good that determines both its fit for the buyer and influences the seller's costs.

	q_ℓ	q_h
θ_h	$U_{h\ell}, C_\ell$	U_{hh}, C_h
θ_ℓ	$U_{\ell\ell}, C_\ell$	$U_{\ell h}, C_h$

TABLE 1. Matrix describing the buyer's value and seller's costs.

(2) *Seller cost monotonicity*: $C_h > C_\ell \geq 0$.
(3) *Efficient trade*: $U_{\ell s} \geq C_s$, for $s \in \{\ell, h\}$.

Assumptions 1 and 2 should be uncontroversial. The first requires the buyer to value the good more when he has a higher type and the good is of higher quality. The second requires higher quality to be more costly. Assumption 3 ensures that the seller's cost is always lower than the buyer's value, that is, trade is efficient. This assumption is not necessary but we impose it nonetheless since doing so simplifies the statements of our results. The economic insights carry over more generally, with more cases to account for regarding the combination of inefficient trade and individual rationality.

Finally, we assume that both players are risk-neutral and that utilities are additively separable and linear in transfers (which we will introduce below).

2.2. *Seller learning*

Why might a seller not perfectly know the quality of her good? The seller of a house or car might simply be unaware or not have the expertise to assess all of the underlying problems or may not know the precise cost of making repairs for issues that she does observe. It may also be that regulatory policies impose restrictions on the information that the seller can use in her pricing decision.

Formally, the seller learns about the quality q of the good via a *binary information structure* $(\Omega, \{G(\cdot|q)\}_{q \in Q})$. Here, $\Omega = \{\omega_h, \omega_\ell\}$, $\omega_h > \omega_\ell$, is a binary set of *signals* and $\{G(\cdot|q)\}_{q \in Q}$ is a pair of conditional distributions, where $g(\omega_s|q) \in [0,1]$ is the probability that signal ω_s, $s \in \{\ell, h\}$ realizes when the quality is $q \in Q$. As with the simplifying assumption on the seller cost, the restriction to binary signals is made to simplify the presentation.[3]

The seller privately observes the signal ω_s and we refer to this as the seller's *type*. She forms a posterior belief $G(\cdot|\omega_s) \in \Delta(Q)$ over the quality of the good using Bayes' rule; we use $g_s := \sum_{s' \in \{\ell,h\}} g(\omega_s|q_{s'}) p_{s'}$ to denote the probability that signal ω_s realizes, for $s \in \{\ell, h\}$.

Observing a signal ω_s allows the seller to form a posterior estimate of her cost $c_s := g(q_h|\omega_s)C_h + g(q_\ell|\omega_s)C_\ell$, for $s \in \{\ell, h\}$. We use $u_{bs} := g(q_h|\omega_s)U_{bh} + g(q_\ell|\omega_s)U_{b\ell}$ to denote the expected value that the buyer has for the good when his type is θ_b and the seller's type is ω_s with $b, s \in \{\ell, h\}$.

Without loss, we assume that signals are labeled so that $g(q_h|\omega_h) \geq g(q_h|\omega_\ell)$ (that is, the good is more likely to be high quality after ω_h is observed). Consequently, $c_h \geq c_\ell$ and $u_{bh} \geq u_{b\ell}$ for both $b \in \{\ell, h\}$.

[3]We have verified that this is without loss when the seller's costs do not depend on the buyer's type and are working to generalize this to the general case discussed in Section 5.

The matrix in Table 2 summarizes the utilities of both players for any given binary information structure.

	ω_ℓ	ω_h
θ_h	$u_{h\ell}, c_\ell$	u_{hh}, c_h
θ_ℓ	$u_{\ell\ell}, c_\ell$	$u_{\ell h}, c_h$

TABLE 2. Matrix describing the buyer's value and seller's costs as a function of their types/signals.

2.3. *The informed principal game*

Since our seller has private information and she proposes the terms of trade, our setting is an informed principal game. Unlike standard mechanism design, different types of the seller could propose distinct mechanisms to the buyer and, consequently, the proposed mechanism also has a signaling component.

The *informed principal game* proceeds as follows where the numerical bullet points denote the timing.

(1) The seller observes signal $\omega \in \Omega$.
(2) The seller proposes a mechanism that consists of
    (a) a finite set of messages for the seller,
    (b) a finite set of messages for the buyer and
    (c) a mapping from messages reported by both players to an allocation (the probability of trade) and a transfer.
(3) The buyer accepts or rejects the mechanism.
(4) If the buyer rejects the mechanism, both players get a payoff of 0.
    If the buyer accepts the mechanism, both players simultaneously pick messages (from their respective message sets) and the allocation and transfer from the mechanism are implemented.

Since we will shortly invoke the inscrutability principle of Myerson (1983), we deliberately choose to be slightly informal and not introduce additional notation to define strategies and equilibrium. In words, the seller's strategy consists of a distribution over finitely many mechanisms followed by a distribution over possible messages should the buyer accept the mechanism (realized from the seller's strategy). The buyer's strategy consists of a decision that determines the probability with which he accepts an offered mechanism followed by a distribution over messages if he accepts the mechanism.

Our solution concept is perfect Bayesian equilibrium, henceforth simply *equilibrium* for short. This requires the following.

(i) The buyer's belief for any on path mechanism is determined by Bayes' rule; beliefs at off path mechanisms are unrestricted.
(ii) Both the buyer's and the seller's strategies are mutual best responses.

(iii) Both the buyer and the seller chooses messages that are mutual best responses at all off path mechanisms (bullet (ii) guarantees mutual best responses at all on path mechanisms).

Note that we impose no further refinements restricting the beliefs of the buyer at off path mechanisms. This choice is deliberate since we want to stack the deck in favor of the seller. We discuss this further after we define a seller-optimal equilibrium.

2.4. *Seller-optimal equilibrium*

A *seller-optimal equilibrium* is an equilibrium that yields the highest seller profit in the set of all equilibria. In this subsection, we define an optimization problem the solution to which allows us to derive a seller-optimal equilibrium.

We first invoke the inscrutability principle of Myerson (1983). This states that it is without loss[4] to assume that *all* seller types choose the *same* incentive compatible and individually rational, *direct* mechanism (in Step 2) which we now define.

A *direct mechanism* is a mechanism where the message spaces for buyer and seller are their respective type spaces. Thus we can write such a mechanism as $(x_{bs}, t_{bs})_{\theta\in\Theta,\omega\in\Omega}$ where the *allocation* $x_{bs} \in [0, 1]$ is the probability that the buyer receives the good and the *transfer* $t_{bs} \in \mathbb{R}$ is the buyer's payment, when the buyer and seller report θ_b and ω_s (with $b, s \in \{\ell, h\}$) respectively.

Since our type space is binary, we can illustrated a direct mechanism in the following matrix (Table 3).

	ω_ℓ	ω_h
θ_h	$x_{h\ell}$, $t_{h\ell}$	x_{hh}, t_{hh}
θ_ℓ	$x_{\ell\ell}$, $t_{\ell\ell}$	$x_{\ell h}$, $t_{\ell h}$

TABLE 3. A direct mechanism: The first value is the allocation and the second is the transfer.

Unlike the typical mechanism design problem (in which the seller has no private information), a direct mechanism in our setting with two-sided private information has two associated sets of incentive compatibility and individual rationality constraints which we define below.

$$\sum_{b\in\{\ell,h\}} f_b(t_{bs} - c_s x_{bs}) \geq \sum_{b\in\{\ell,h\}} f_b(t_{bs'} - c_{s'} x_{bs'}) \qquad \forall s, s' \in \{\ell, h\} \tag{IC_S}$$

$$\sum_{b\in\{\ell,h\}} f_b(t_{bs} - c_s x_{bs}) \geq 0 \qquad \forall s \in \{\ell, h\}, \tag{IR_S}$$

$$\sum_{s\in\{\ell,h\}} (u_{bs} x_{bs} - t_{bs})\, g_s \geq \sum_{s\in\{\ell,h\}} (u_{bs} x_{b's} - t_{b's}) g_s \quad \forall b, b' \in \{\ell, h\}, \tag{IC_B}$$

$$\sum_{s\in\{\ell,h\}} (u_{bs} x_{bs} - t_{bs})\, g_s \geq 0 \qquad \forall b \in \{\ell, h\}. \tag{IR_B}$$

The individual rationality constraints (IR_S, IR_B for the seller, buyer respectively) ensure that each player receives a non-negative utility. The incentive compatibility constraints (IC_S, IC_B for the

[4]"Without loss" here means that, given any equilibrium, there exists another equilibrium in which all seller types offer the same mechanism such that allocations and transfers for every pair of types (θ_b, ω_s) is the same in both equilibria.

seller, buyer respectively) ensure that each player reports their type truthfully given their opponent is reporting truthfully as well.

Because the inscrutability principal is perhaps less familiar than the standard revelation principle (without seller private information), we provide some intuition for why it is without loss to assume that all seller types offer the same mechanism. Consider an equilibrium in which different sellers offer different (potentially non incentive compatible and individually rational) mechanisms. Suppose instead that all seller types offer the same direct mechanism in which, for every pair of types (θ_b, ω_s), the expected allocation and transfer (x_{bs}, t_{bs}) correspond precisely to the equilibrium outcome. For simplicity, suppose the buyer's belief at any off path mechanism assigns probability 1 to the worse quality seller, type ω_ℓ.

This mechanism will be incentive compatible and individually rational for both players; thus, it will be an equilibrium for all seller types to offer this mechanism (since their on path utility is the same and the buyer has the worst possible off path belief). Loosely speaking, the former holds because, in the equilibrium, it is not optimal for a seller of a given type ω_s to deviate and to offer the mechanism chosen by a different type $\omega_{s'}$ and then follow the message reporting strategy of $\omega_{s'}$ in Step 4 (and thereby obtain exactly the payoff from misreporting as $\omega_{s'}$ in the direct mechanism). In fact, the direct mechanism shrinks the set of possible deviations for the seller (because type ω_s could choose a different reporting strategy to $\omega_{s'}$ in Step 4). Conversely, the buyer gets no information about the seller's type when all types offer the same direct mechanism. He will thus not deviate because he receives the same payoff as in the given equilibrium and has less information. Individual rationality follows from the fact that both players, irrespective of type, can always guarantee themselves a payoff of zero: the seller by choosing not to sell the good at transfer zero and the buyer by rejecting the mechanism in Step 3.

Now, let

$$\pi_s := \max\{f_h[u_{h\ell} - c_s], u_{\ell\ell} - c_s, 0\}$$

denote the highest possible profit that a seller of type ω_s can obtain if the buyer's belief assigns probability one to the seller's type being ω_ℓ. With this last piece of notation in place we can set up the following linear program that characterizes the seller-optimal equilibrium.

$$\begin{aligned} &\max_{(x,t)} \sum_{b,s\in\{\ell,h\}} g_s f_b (t_{bs} - c_s x_{bs}) \\ &\text{subject to} \\ &IC_S,\ IC_B,\ IR_B \text{ and} \\ &\sum_{b\in\{\ell,h\}} f_b(t_{bs} - c_s x_{bs}) \geq \pi_s \qquad \forall s \in \{\ell, h\}. \end{aligned} \tag{1}$$

The objective function in (1) is the seller's expected profit when all types of the seller offer the direct mechanism (x, t). The first three constraints ensure that the buyer accepts the mechanism (a consequence of the constraint IR_B) and that, after the buyer accepts the mechanism, both the seller and the buyer report their types truthfully (consequences of the constraints IC_S and IC_B). The last constraint (which is a stronger constraint than IR_B) ensures that the seller has no incentive to deviate when the buyer's off path belief assigns probability one to the seller being type ω_ℓ. Note that any equilibrium can be supported with this off path buyer belief since either seller type can obtain a weakly higher profit for any other buyer belief.

In what follows we informally refer to a solution of the optimization problem (1) (instead of a pair of strategies and off path buyer beliefs) as a seller-optimal equilibrium.

2.5. *Buyer-Optimal Outcome*

Our goal is find the information structure and the corresponding seller-optimal equilibrium that maximize consumer surplus.

We refer to $\{(\Omega, \{G(\cdot|q)\}_{q\in Q}), (x,t)\}$ as an *outcome*, whenever

- $(\Omega, \{G(\cdot|q)\}_{q\in Q})$ is a binary information structure via which the seller learns the quality and
- given this information structure, (x,t) is a solution to (1) (and is thus a seller-optimal equilibrium).

The *buyer-optimal outcome* $\{(\Omega, \{G^*(\cdot|q)\}_{q\in Q}), (x^*,t^*)\}$ maximizes consumer surplus across all outcomes. Note that, implicit in this definition is the fact that, when there are multiple seller-optimal equilibria, the buyer-optimal outcome selects the one that maximizes the consumer surplus.

We end this section with justification for our focus on the buyer-optimal outcome. Inherent in this are two choices: the first is to focus on consumer surplus and the second is to restrict attention to seller-optimal equilibria. We have already motivated the first of these two in the introduction but we reiterate here briefly. From a theoretical perspective, one might expect that additional seller information always yields higher profits and, in turn, lower consumer surplus (neither of these are true). From an applied perspective, our analysis aims to inform the discussion of consumer privacy. Indeed we show that, under certain circumstances, seller private information may lead to Pareto improvements.

The seller-optimal equilibrium is a natural candidate equilibrium to study for a seller who wants to maximize profits. In deriving this equilibrium, we obtain the solution of the informed principal optimum for a canonical two-type interdependent value setting. It was surprising to us that this had not previously been derived and consequently, we view this to be one of our significant contributions. Second, we felt that our insight—that seller private information can lead to greater consumer surplus—would be starker and more counterintuitive when the seller is allowed to use her private information to maximize profits.

3. THE SELLER-OPTIMAL EQUILIBRIUM

In this section, we derive the seller-optimal equilibrium. But first, we discuss a benchmark where the seller has no information. The seller-optimal equilibrium is trivial in this case but it motivates the parameters of interest.

3.1. *Benchmark: No information*

We first consider the case where the seller has no information about the quality; this corresponds to an information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$ in which

$$g(\omega_h|q_h) = g(\omega_h|q_\ell) = g(\omega_\ell|q_h) = g(\omega_\ell|q_\ell) = \frac{1}{2} \tag{2}$$

When the seller has no information, this is effectively a standard monopoly problem where the buyer has two types. It is optimal for the seller to choose one of two prices: the expected value of the high-type buyer or the expected value of the low-type buyer. The latter is optimal whenever

$$\begin{aligned} & p_h(U_{\ell h} - C_h) + p_\ell(U_{\ell\ell} - C_\ell) \geq f_h\left[p_h(U_{hh} - C_h) + p_\ell(U_{h\ell} - C_\ell)\right] \\ \iff \quad & p_h \underbrace{(U_{\ell h} - U_{hh} f_h - C_h f_\ell)}_{:=\varphi_h} + p_\ell \underbrace{(U_{\ell\ell} - U_{h\ell} f_h - C_\ell f_\ell)}_{:=\varphi_\ell} \geq 0. \end{aligned}$$

In terms of our notation, this corresponds to the seller-optimal equilibrium (x, t) where

$$x_{hh} = x_{h\ell} = x_{\ell h} = x_{\ell\ell} = 1 \text{ and } t_{hh} = t_{h\ell} = t_{\ell h} = t_{\ell\ell} = p_h U_{\ell h} + p_\ell U_{\ell\ell}. \tag{3}$$

In words, the seller sells to both buyer types at a price equal to the expected value of the low-type buyer.

The shorthand notation φ_ℓ and φ_h defined below the underbraces refers to the standard "virtual values" of a seller whose good is commonly known to be low- and high-quality respectively. The sign of φ_s determines whether a monopolist whose good is commonly known to be value q_s finds it optimal to sell to both buyer types ($\varphi_s \geq 0$) or just to the high-type buyer ($\varphi_s < 0$).

We now observe that when $p_h\varphi_h + p_\ell\varphi_\ell \geq 0$, the buyer-optimal outcome is immediate: the seller receives no information (as in 2) and the seller-optimal equilibrium involves setting a price equal to the expected value of the low-type buyer (as in 3).

To see this, examine the optimization problem (1) corresponding to the seller-optimal equilibrium. As in standard mechanism design problems, and for the same reasons, the low-type buyer's individual rationality constraint IR_B and the high to low type buyer's incentive compatibility constraint must bind. If IR_B was slack for the low-type buyer, the seller could just increase all transfers t_{bs} by the same constant amount without violating any of the buyer's or seller's constraints and get a higher profit. Similarly, if the high to low type buyer's IC_B constraint was slack, the seller could raise the transfers of t_{hs} by the same constant amount and, once again, get a higher profit because this change would not violate any of the buyer's or seller's constraints.

Therefore, the low-type buyer always gets utility zero in a seller optimal equilibrium and consequently, consumer surplus accrues to the buyer only in the form of information rents that the seller must surrender to the high-type buyer. These rents are increasing in the allocation to the low-type buyer and therefore, the highest possible consumer surplus is achieved whenever the low-type buyer receives the good with probability one which is precisely what happens in the proposed buyer-optimal outcome.

Consequently, in what follows, we restrict attention to the case of $p_h\varphi_h + p_\ell\varphi_\ell < 0$.

3.2. *Characterization*

We now consider the case where the seller learns the quality of the good via a binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$.

We first define

$$\phi_s := g(q_h|\omega)\varphi_h + g(q_\ell|\omega)\varphi_\ell$$

to be the virtual value for the seller type ω and note that

$$p_\ell\varphi_\ell + p_h\varphi_h = g_\ell\phi_\ell + g_h\phi_h < 0.$$

Our first result characterizes the seller-optimal equilibrium.

THEOREM 1. *Suppose that $p_\ell\varphi_\ell + p_h\varphi_h < 0$. The allocation rule of every seller-optimal equilibrium (x,t) takes the form*

$$x_{h\ell} = 1, \quad x_{hh} \in (0,1] \quad \textit{and} \quad x_{\ell\ell} = x_{\ell h} = 0,$$

if $\phi_\ell < 0$, $\phi_h \geq 0$, and

$$x_{h\ell} = x_{\ell\ell} = 1, \quad x_{hh} \in (0,1] \quad \textit{and} \quad x_{\ell h} = 0,$$

if $\phi_\ell > 0$, $\phi_h < 0$.
When $\phi_\ell = 0$, $\phi_h < 0$, there is a seller-optimal equilibrium of the above form (but there may be others in which $x_{\ell h} = 0$ but $x_{\ell\ell} \in [0,1]$).

The proof of the more general result Theorem 3 (which incorporates Theorem 1 as a special case) is provided in the appendix.

Given the equilibrium allocation characterized in Theorem 1, the expected transfer ($\sum_{s\in\{\ell,h\}} g_s t_{bs}$) for each buyer type θ is pinned down by the two binding constraints for the buyer: the high to low type IC_B constraint and the low-type IR_B constraint or

$$\sum_{s\in\{\ell,h\}} (u_{hs}x_{hs} - t_{hs})\, g_s = \sum_{s\in\{\ell,h\}} (u_{hs}x_{\ell s} - t_{\ell s}) g_s \text{ and } \sum_{s\in\{\ell,h\}} (u_{\ell s}x_{\ell s} - t_{\ell s})\, g_s = 0$$

respectively. It is typically possible to decompose these expected transfers in more than one way to ensure seller incentive compatibility and that neither seller type has an incentive to deviate.

An implication of Theorem 1 is that, when $\phi_\ell < 0$ and $\phi_h \geq 0$, the buyer receives zero consumer surplus.[5] Consequently, the buyer is indifferent between the seller having no information and her learning via a binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$. Conversely, when $\phi_\ell > 0$ and $\phi_h \leq 0$, the buyer receives positive consumer surplus (because $x_{\ell\ell} = 1$) and, therefore, he strictly prefers the seller to have private information over having no information!

In this second case, why does seller private information benefit the buyer? When the seller has no information, it is not profitable for her to sell to the low type buyer because $p_\ell\varphi_\ell + p_h\varphi_h < 0$. When $\phi_\ell > 0$, the low-type seller gets a higher profit by selling to both buyer types at a price $g(q_h|\omega_\ell)u_{\ell h} + g(q_\ell|\omega_\ell)u_{\ell\ell}$ as opposed to selling to only the high-type buyer at a price of $g(q_h|\omega_\ell)u_{hh} + g(q_\ell|\omega_\ell)u_{h\ell}$. Because $\phi_h < 0$, the converse is true for the high-type seller. Because

[5] We show that when the seller's costs depend on the buyer's type, then there are conditions under which the buyer receives positive surplus, even in the case $\phi_h > 0$.

the seller has information about the quality, Theorem 1 shows that it possible, and profitable, to implement an allocation in which both seller types allocate differently to the buyer.

Why does a similar logic not apply fo the first case of $\phi_\ell < 0$ and $\phi_h \geq 0$? The answer lies in the seller's incentive compatibility constraint IC_S. As in standard mechanism design, incentive compatibility implies monotonicity. Since we have incentive constraints for both players, this in particular implies that the low (cost) type seller must have a higher expected allocation or that $f_h x_{h\ell} + f_\ell x_{\ell\ell} \geq f_h x_{hh} + f_\ell x_{\ell h}$. Consequently, there exist no transfers that can make an allocation of the form $x_{h\ell} = x_{hh} = x_{\ell h} = 1$ and $x_{\ell\ell} = 0$ incentive compatible for the seller. Thus, if the high-type seller wants to allocate to both buyer types (so $x_{hh} = x_{\ell h} = 1$), incentive compatibility implies that the low-type seller must also allocate to both buyer types (so $x_{h\ell} = x_{\ell\ell} = 1$). But this cannot be a seller-optimal equilibrium because $p_\ell \varphi_{q_\ell} + p_h \varphi_h < 0$ implies that it is not optimal to always sell the good.

Lastly, why is x_{hh} sometimes interior? The reason is that the transfers required to make an allocation of $x_{hh} = 1$ incentive compatible might cause a violation of the seller's equilibrium deviation constraint (the last constraint in the linear program 1). We now demonstrate the above intuition via an example.

Example. In this example we compare the benchmark of no information to full information with the intention of showing that the latter can lead to higher consumer surplus. Suppose both buyer types and seller qualities are equally likely ($f_h = f_h = p_h = p_\ell = \frac{1}{2}$). Consider the following utilities and costs for the buyer and seller respectively (Table 4):

	q_ℓ	q_h
θ_h	3, 0	14, 8
θ_ℓ	2, 0	10, 8

Table 4. Example: The buyer's value U_{bs} and seller's cost C_s.

Note that these values are such that the signs of $\varphi_h = U_{\ell h} - U_{hh} f_h - c_h f_\ell = -1$, $\varphi_\ell = U_{\ell\ell} - U_{h\ell} f_h - c_\ell f_\ell = .5$ and $p_h \varphi_h + p_\ell \varphi_\ell = -.25$ satisfy the conditions of the second statement of Theorem 1.

So, first, consider the information structure $(\Omega, \{G(\cdot|q)\}_{q \in Q})$ with

$$g(\omega_h|q_h) = g(\omega_h|q_\ell) = g(\omega_\ell|q_h) = g(\omega_\ell|q_\ell) = \frac{1}{2}$$

that provides the seller with no information about her type. Table 5 describes the resulting type space and the seller-optimal equilibrium.

Here trade happens with probability $\frac{1}{2}$, the seller's expected profit is $\sum_{b,s \in \{\ell,h\}} g_s f_b (t_{bs} - c_s x_{bs}) = \frac{1}{2} \times (8.5 - 4) = 2.25$ and the buyer's surplus is $\sum_{b,s \in \{\ell,h\}} g_s f_b \,(u_{bs} x_{bs} - t_{bs}) = 0$.

Now consider the perfectly informative information structure $(\Omega, \{G(\cdot|q)\}_{q \in Q})$ with $g(\omega_h|q_h) = g(\omega_\ell|q_\ell) = 1$ and $g(\omega_h|q_\ell) = g(\omega_\ell|q_h) = 0$. Table 6 describes the resulting type space and the seller-optimal equilibrium:

It is easy to verify that the seller-optimal equilibrium satisfies all the constraints of (1). The low to high seller IC_S constraint binds (same expected transfer and same cost from misreporting)

payoffs			*seller-optimal equilibrium*		
	ω_ℓ	ω_h		ω_ℓ	ω_h
θ_h	8.5, 4	8.5, 4	θ_h	1, 8.5	1, 8.5
θ_ℓ	6, 4	6, 4	θ_ℓ	0, 0	0, 0

TABLE 5. No information: Buyer's value u_{bs}, seller's cost c_s, allocation x and transfer t in the seller-optimal equilibrium.

payoffs			*seller-optimal equilibrium*		
	ω_ℓ	ω_h		ω_ℓ	ω_h
θ_h	3, 0	14, 8	θ_h	1, 8	1, 8
θ_ℓ	2, 0	10, 8	θ_ℓ	1, 1	0, 1

TABLE 6. Full information: Buyer's value u_{bs}, seller's cost c_s, allocation x and transfer t in the seller-optimal equilibrium.

and the high to low type IC_S constraint is slack (same expected transfer but higher costs from misreporting). The low-type buyer's IR_B constraint binds as does the high to low buyer IC_B constraint (the low to high IC_B constraint is slack). Finally, the highest profits that the seller can achieve from deviating are $\pi_h = 0$ and $\pi_\ell = 2$ which are lower than their profits in the seller optimal equilibrium.

Trade now happens with probability $\frac{3}{4}$, the seller's expected profit is $\sum_{b,s\in\{\ell,h\}} g_s f_b(t_{bs} - c_s x_{bs}) = \frac{1}{2} \times (8-4) + \frac{1}{2} \times (1-0) = 4.5$ and the buyer's surplus is $\sum_{b,s\in\{\ell,h\}} g_s f_b \left(u_{bs}x_{bs} - t_{bs}\right) = \frac{1}{2} \times (8.5 - 8) = .25$. In other words, seller private information leads to a Pareto improvement (relative to no information)!

The source of these Pareto gains is the ability of different types of the privately informed seller to profitably offer distinct allocations that satisfy the appropriate monotonicity properties (the low-type seller's and the high-type buyer's allocation is higher) to obtain incentive compatibility. In this example, the solution is not interior because the seller's deviation constraint is slack.

Suppose we change the example such that the cost of the high-type seller is now 10 (this does not alter the signs of φ_h, φ_ℓ and $p_h\varphi_h + p_\ell\varphi_\ell$) and we consider the same candidate seller-optimal equilibrium (the transfers were derived from the buyer's binding IC_B and IR_B constraints so did not depend on seller costs). This is illustrated in Table 7

Note that now the high-type seller's profit is less than zero so this cannot be an equilibrium since she will deviate. Observe that, since the low-type seller's cost is zero, seller IC_S implies that the expected transfer to the low-type must be higher. In other words, with allocation, there are no transfers that satisfy seller IC_S that additionally ensure that the high-type seller is not better off deviating. So, in this case, the seller optimal equilibrium will be interior as presented in Table 8.

	payoffs		*seller-optimal equilibrium*		
	ω_ℓ	ω_h		ω_ℓ	ω_h
θ_h	$3, 0$	$14, 10$	θ_h	$1, 8$	$1, 8$
θ_ℓ	$2, 0$	$10, 10$	θ_ℓ	$1, 1$	$0, 1$

TABLE 7. Full information: Buyer's value u_{bs}, seller's cost c_s, allocation x and transfer t in a candidate equilibrium.

	payoffs		*seller-optimal equilibrium*		
	ω_ℓ	ω_h		ω_ℓ	ω_h
θ_h	$3, 0$	$14, 10$	θ_h	$1, 5\frac{1}{3}$	$\frac{2}{3}, 6\frac{2}{3}$
θ_ℓ	$2, 0$	$10, 10$	θ_ℓ	$1, 2$	$0, 0$

TABLE 8. Full information: Buyer's value u_{bs}, seller's cost c_s, allocation x and transfer t in the seller-optimal equilibrium.

4. THE BUYER-OPTIMAL OUTCOME

In this section, we characterize the buyer-optimal outcome. Theorem 1 characterizes the seller-optimal equilibrium corresponding to a given information structure. So, it remains to characterize the information structure in the buyer-optimal outcome which we do next.

THEOREM 2. *Suppose that $p_\ell\varphi_\ell + p_h\varphi_h < 0$.*

If $\varphi_\ell < 0$, $\varphi_h \geq 0$, the buyer gets zero consumer surplus in every outcome. Consequently, every outcome is buyer optimal.

If $\varphi_\ell \geq 0$, $\varphi_h < 0$, the binary information structure $(\Omega, \{G^*(\cdot|q)\}_{q\in Q})$ in the buyer-optimal outcome is given by

$$g^*(w_\ell|q_\ell) = 1, \text{ and } g^*(\omega_\ell|q_h) = -\frac{p_\ell\varphi_\ell}{p_h\varphi_h}$$

with $g^*(w_h|q_s) = 1 - g^*(w_\ell|q_s)$ for $s \in \{\ell, h\}$. The corresponding seller-optimal equilibrium (x^*, t^*) for this information structure (as characterized in Theorem 1) has the property that $x^*_{\ell\ell} = 1$.

The theorem answers the question we posed in the title. When $\varphi_\ell < 0$, $\varphi_h \geq 0$, the seller only sells to the high-type buyer irrespective of information structure and extracts all the surplus from trade. Consequently, the buyer is indifferent between all outcomes and, in particular, seller private information does not hurt the buyer (but it may hurt the seller).

Conversely, when $\varphi_\ell \geq 0$, $\varphi_h < 0$, the low-type seller would trade with the low-type buyer under the information structure that allows the seller to perfectly learn the quality (apply Theorem 1) leading to positive consumer surplus; therefore, the buyer-optimal outcome must feature positive consumer surplus. Indeed, when seller private information benefits the buyer, such information is interior except for the knife-edge case of $\varphi_\ell = 0$. The buyer-optimal information structure is such that ω_h perfectly reveals the quality to be q_h but ω_ℓ does not imply the seller's type is surely

q_ℓ. To summarize our main insight, *whenever trade is inefficient with no seller private information, the buyer (weakly) prefers the seller to be privately informed.*

We now provide some intuition. When $\varphi_\ell < 0$ and $\varphi_h \geq 0$, then under every binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$, the seller-optimal equilibrium (x,t) has the feature that $x_{\ell\ell} = x_{\ell h} = 0$. To see this, first observe that

$$\phi_\ell = g(q_h|\omega_\ell)\varphi_h + g(q_\ell|\omega_\ell)\varphi_\ell \leq g(q_h|\omega_h)\varphi_h + g(q_\ell|\omega_h)\varphi_\ell = \phi_h$$

because $g(q_h|\omega_\ell) \leq g(q_h|\omega_h)$ and $\varphi_h > \varphi_\ell$ (by assumption). Consequently, it is not possible to have $\phi_\ell \geq 0$ since this will imply that $g_\ell\phi_\ell + g_h\phi_h = p_\ell\varphi_\ell + p_h\varphi_h \geq 0$ which is a contradiction. Therefore, for any binary information structure $\phi_\ell < 0$, and Theorem 1 then implies that the buyer gets zero consumer surplus.

When $\varphi_\ell > 0$ and $\varphi_h < 0$ (we ignore the knife-edge case of $\varphi_\ell = 0$ in this discussion), then there is at least one binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$—the seller perfectly learns about the quality—with the feature that $\phi_\ell > 0$. Now, consider any information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$ for which $\phi_\ell > 0$ (and consequently $\phi_h < 0$). The consumer surplus is

$$\begin{aligned}
\sum_{b,s\in\{\ell,h\}} g_s f_b\left(u_{bs}x_{bs} - t_{bs}\right) &= g_\ell f_h(u_{h\ell} - u_{\ell\ell}) \\
&= g_\ell f_h[(u_{hh} - u_{\ell h})g(q_h|\omega_\ell) + (u_{h\ell} - u_{\ell\ell})g(q_\ell|\omega_\ell)] \\
&= f_h(u_{hh} - u_{\ell h})g(\omega_\ell|q_h)p_h + f_h(u_{h\ell} - u_{\ell\ell})g(\omega_\ell|q_\ell)p_\ell.
\end{aligned}$$

The highest consumer surplus is obtained by choosing $g(\omega_\ell|q_h)$ and $g(\omega_\ell|q_\ell)$ to maximize the above expression subject to the constraint

$$\begin{aligned}
&\phi_\ell = g(q_h|\omega_\ell)\varphi_h + g(q_\ell|\omega_\ell)\varphi_\ell \geq 0 \\
\Longleftrightarrow &g(\omega_\ell|q_h)p_h\varphi_h + g(\omega_\ell|q_\ell)p_\ell\varphi_\ell \geq 0,
\end{aligned}$$

which is the condition that ensures $x_{\ell\ell} = 1$ in a seller-optimal equilibrium. Since $\varphi_\ell > 0$, $\varphi_h < 0$, $u_{hh} - u_{\ell h} > 0$ and $u_{h\ell} - u_{\ell\ell} > 0$, there is a buyer-optimal outcome with the information structure

$$g^*(\omega_\ell|q_\ell) = 1 \text{ and } g^*(\omega_\ell|q_h) = -g^*(\omega_\ell|q_\ell)\frac{p_\ell\varphi_\ell}{p_h\varphi_h} = -\frac{p_\ell\varphi_\ell}{p_h\varphi_h}.$$

Note that $\phi_\ell = 0$ implies that the low-type seller is indifferent between selling to both types (by charging a price $u_{\ell\ell}$) and just to the high-type buyer (by charging a price $u_{h\ell}$). An implication is that, if $x^*_{hh} = 1$ in the buyer-optimal outcome, the seller's profit is the same as what she would get if she received no information whatsoever. Consequently, in this case, the buyer-optimal outcome has the feature that seller private information leads to gains in total surplus that arise because of the increased probability of trade (from $x^*_{\ell\ell} = 1$) but all those gain accrue to the buyer! When $x^*_{hh} < 1$ in the buyer-optimal outcome, the seller's profit is lower than what she would receive absent any information: $g_h u_{hh} x^*_{hh} + g_\ell u_{h\ell}$ versus $g_h u_{hh} + g_\ell u_{h\ell} = p_h u_{hh} + p_\ell u_{h\ell}$. Perhaps counter-intuitively, in this case, seller-private information hurts profits but boosts consumer surplus.

We end this section by revisiting the example.

EXAMPLE (CONTINUED). Recall that both buyer types and seller qualities are equally likely, and that the utilities and costs for the buyer and seller respectively are the following (Table 9).

	q_ℓ	q_h
θ_h	$3, 0$	$14, 8$
θ_ℓ	$2, 0$	$10, 8$

TABLE 9. Example: The buyer's value U_{bs} and seller's cost C_s.

Here, $\varphi_h = u_{\ell h} - u_{hh} f_h - c_h f_\ell = -1$ and $\varphi_\ell = u_{\ell\ell} - u_{h\ell} f_h - c_\ell f_\ell = .5$. Theorem 2 states that the buyer-optimal outcome has the information structure

$$g^*(w_\ell | q_\ell) = 1, \text{ and } g^*(\omega_\ell | q_h) = -\frac{p_\ell \varphi_\ell}{p_h \varphi_h} = \frac{1}{2}.$$

The resulting type space and the seller-optimal equilibrium is the following (Table 10):

payoffs

	ω_ℓ	ω_h
θ_h	$6\frac{2}{3}, 2\frac{2}{3}$	$14, 8$
θ_ℓ	$4\frac{2}{3}, 2\frac{2}{3}$	$10, 8$

$g_\ell = \frac{3}{4} \quad g_h = \frac{1}{4}$

seller-optimal equilibrium

	ω_ℓ	ω_h
θ_h	$1, 6\frac{2}{3}$	$1, 8$
θ_ℓ	$1, 4\frac{2}{3}$	$0, 0$

TABLE 10. Full information: Buyer's value u_{bs}, seller's cost c_s, allocation x^* and transfer t^* in the seller-optimal equilibrium corresponding to the buyer-optimal outcome.

Trade happens with probability $\frac{7}{8}$, the seller's expected profit is $\sum_{b,s \in \{\ell,h\}} g^*_s f_b (t_{bs} - c_s x_{bs}) = \frac{3}{4} \times 3 + \frac{1}{4} \times 0 = 2.25$ and the buyer's surplus is $\sum_{b,s \in \{\ell,h\}} g^*_s f_b (u_{bs} x_{bs} - t_{bs}) = \frac{1}{2} \times 1.5 + \frac{1}{2} \times 0 = .75$. Note that the buyer's surplus is higher than what he would receive (the previously computed value of .25) if the seller was fully informed. Conversely, the seller's profit is identical to what she gets with no information. Here, all additional gains from trade that can be realized due to seller private information accrue to the buyer as information rents.

5. DISCUSSION

5.1. *Generalizing the seller's costs*

One of our motivating applications in the introduction was that of health insurance. Here, as we discussed, the seller's cost can depend both on the quality and the buyer's type. In this subsection, we demonstrate how our results extend to this fully interdependent value case.

The seller's cost C_{bs} now has the additional subscript b that captures the dependence on the buyer's type. The assumptions on the seller's cost remain unchanged:

(2′) *Seller cost monotonicity*: $C_{bh} > C_{b\ell} \geq 0$ for $b \in \{\ell, h\}$.

(3′) *Efficient trade*: $U_{bs} \geq C_{bs}$, for $b, s \in \{\ell, h\}$.

Note that, once again, the seller's cost increases in quality for each buyer type and that trade is always efficient. Note that we are not restricting how the buyer's type θ_b affect the seller's cost for a given quality q_s; we maintain this generality since it does not complicate the presentation.

The remaining notation generalizes similarly. The virtual values

$$\varphi_h = U_{\ell h} - U_{hh} f_h - C_{\ell h} f_\ell,$$
$$\varphi_\ell = U_{\ell\ell} - U_{h\ell} f_h - C_{\ell\ell} f_\ell$$

are defined as before with costs $C_{\ell s}$, $s \in \{\ell, h\}$ corresponding to the low-type buyer. The notation overload is deliberate as there should be no confusion since the meaning is clear in this subsection. For any binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$, the seller's posterior estimate of her cost is $c_{bs} := g(q_h|\omega_s)C_{bh} + g(q_\ell|\omega_s)C_{b\ell}$ when she observes signal ω_s and the buyer's type is θ_b. The virtual values ϕ_s are analogously defined.

We are now in a position to describe the seller-optimal equilibrium corresponding to an arbitrary binary information structure.

THEOREM 3. *Suppose that* $p_\ell \varphi_\ell + p_h \varphi_h < 0$.

When $\phi_\ell < 0$, $\phi_h > 0$ *and* $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\phi_h \le f_h(u_{hh}-c_{hh})$, *the allocation rule of every seller-optimal equilibrium* (x,t) *takes the form*

$$x_{h\ell} = 1, \;\; x_{hh} \in (0,1] \;\; \textit{and} \;\; x_{\ell\ell} = x_{\ell h} = 0.$$

When $\phi_\ell < 0$, $\phi_h > 0$ *and* $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\phi_h > f_h(u_{hh}-c_{hh})$, *the allocation rule of every seller-optimal equilibrium* (x,t) *takes the form*

$$\begin{aligned} &x_{h\ell} = 1, \;\; x_{hh} = x_{\ell\ell} = 0 \;\; \textit{and} \;\; x_{\ell h} \in (0,1] && \textit{if} \;\; \tfrac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})} < 1, \\ &x_{h\ell} = 1, \;\; x_{hh} \in (0,1], \;\; x_{\ell\ell} = 0 \;\; \textit{and} \;\; x_{\ell h} = 1 && \textit{if} \;\; \tfrac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})} > 1. \end{aligned}$$

When $\phi_\ell > 0$, $\phi_h < 0$, *the allocation rule of every seller-optimal equilibrium* (x,t) *takes the form*

$$x_{h\ell} = x_{\ell\ell} = 1, \;\; x_{hh} \in [0,1] \;\; \textit{and} \;\; x_{\ell h} = 0.$$

While the above statement somewhat laboriously categorizes all the possible cases, the main content is easily summarized. When $\phi_\ell > 0$, $\phi_h < 0$ (the last case), the allocation in the seller-optimal equilibrium takes the same form as in the case where the seller's costs depended on quality alone (Theorem 1). However, recall that when $\phi_\ell < 0$, $\phi_h > 0$, the low-type buyer was never allocated the good with positive probability in Theorem 1.

This is no longer the case when the seller's costs are interdependent and $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\phi_h > f_h(u_{hh} - c_{hh})$; in this case the low-type buyer is allocated the good with positive probability when the seller receives the high signal ω_h. An implication is that, with interdependent costs, there are a broader set of parameters under which the buyer would strictly prefer the seller to privately know the quality over having no information at all.

For intuition, note that with interdependent costs, it is possible to have $f_\ell(u_{\ell h} - c_{\ell h}) > f_h(u_{hh} - c_{hh})$ which implies that the high-type seller prefers to sell only to the low-type buyer as opposed to selling only to the high-type buyer. Moreover, if $(u_{\ell h} - c_{hh}) < 0$, the high-type seller would strictly prefer to sell only to the low-type buyer as opposed to charging a price $u_{\ell h}$ and selling to both buyer types. Note that neither of these are possible when costs depend solely on quality.

If the seller's type was publicly known to be ω_h (as in standard monopoly pricing), IC_B makes it impossible to allocate to the low-type buyer with a higher probability than to the high-type. But when the seller has private information, IC_B only requires the buyer's allocation to be monotone in expectation and, it is this flexibility that allows the seller to offer mechanisms such as those in Theorem 3. But of course, allocations must also respect IC_S and hence the condition in Theorem 3 is not as transparent as this high-level intuition might suggest.

As it is tedious, we do not spell out the buyer-optimal outcome as we did in Theorem 2. Instead, we use the characterization of the seller-optimal equilibrium to emphasize that the two key insights from the benchmark model generalize: (i) seller private information can lead to higher consumer surplus and (ii) in this case, the binary information structure in the buyer-optimal outcome is typically noisy.

The first of these two insights follows immediately from Theorem 3 and so we end this section by discussing the second. When $\phi_\ell > 0$, $\phi_h < 0$, the binary information structure in the buyer-optimal outcome is identical to that in Theorem 2. The difference is when $\phi_\ell < 0$, $\phi_h > 0$; but here too a similar intuition applies. For simplicity, consider the case where $\frac{f_h(C_{hh}-C_{h\ell})}{f_\ell(C_{\ell h}-C_{\ell\ell})}\varphi_h > f_h(U_{hh} - C_{hh})$ and $\frac{f_h(C_{hh}-C_{h\ell})}{f_\ell(C_{\ell h}-C_{\ell\ell})} > 1$ which has the property that $x_{\ell h} = 1$ in the seller-optimal equilibrium with full information. Recall that the source of consumer surplus is the extent to which the low-type buyer is allocated the good.

So, for $\varepsilon \in (0,1)$, consider a binary information structure $(\Omega, \{G(\cdot|q)\}_{q\in Q})$ in which

$$g(w_\ell|q_\ell) = 1-\varepsilon, \text{ and } g(\omega_h|q_h) = 1.$$

In words, the signal w_ℓ perfectly reveals the quality to be q_ℓ but the posterior following ω_h assigns positive probability to the quality being q_ℓ. For sufficiently small ε, we will have $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\phi_h > f_h(u_{hh} - c_{hh})$ and $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})} > 1$ and so the seller-optimal equilibrium will have $x_{\ell h} = 1$. Since the low-type buyer is now being allocated the good with greater probability, this outcome must yield strictly higher consumer surplus than when the seller has full information. Consequently, the buyer-optimal outcome must feature a binary information structure that is noisy.

5.2. *Implementing the seller-optimal equilibrium*

As in all mechanism design papers, the direct mechanism that describes the seller's on-path behavior is a theoretical object that simplifies solving for the seller-optimal equilibrium. In this subsection, we discuss how such mechanisms can be implemented in practice. For clarity of presentation, we demonstrate this via an example but our suggested implementation generalizes. Here, we take the (equivalent) interpretation that there is not a single buyer but a mass one of buyers of whom a fraction f_b have type θ_b, $b \in \{\ell, h\}$.

Now, consider the following game.

(1) The seller first offers a menu of *membership* options. Each option consists of an up-front membership fee and a pair of price-contingent discounts (which we define more precisely below).
(2) Each buyer chooses an option from the menu (or can refuse to participate) and pays the up-front fee.
(3) The seller chooses a price to offer to all the buyers.
(4) The buyer decides whether or not to purchase the good at the discounted offered price: this is the offered price minus the buyer's discount given by the membership option he purchased.

Specifically, each membership option (targeted for the buyer of type θ_b) takes the form $(m_b, \beta_b, \underline{d}_b, \overline{d}_b)$ where m_b is the up-front membership fee, β_b is the threshold and $\underline{d}_b$, $\overline{d}_b$ are the discounts if the seller offers a price below or (weakly) above the threshold respectively.

Consider the following example with payoffs and the corresponding seller-optimal equilibrium as follows (Table 11) – all types are equally likely:

payoffs	ω_ℓ	ω_h		*seller-optimal equilibrium*	ω_ℓ	ω_h
θ_h	3, 0	10, 1		θ_h	1, 6	1, 6
θ_ℓ	2, 0	4, 1		θ_ℓ	1, 1	0, 1

TABLE 11. Full information: Buyer's value u_{bs}, seller's cost c_s, allocation x and transfer t in the seller-optimal equilibrium.

Note that the profit that seller-type ω_s receives from reporting types ω_ℓ and ω_h is $\frac{1}{2} \times 6 + \frac{1}{2}1 - c_s = 3.5 - c_s$ and $\frac{1}{2} \times 6 + \frac{1}{2}1 - \frac{1}{2} \times c_s = 3.5 - \frac{c_s}{2}$ respectively.

Now consider the following strategies. Both seller types first offer the menu $(m_h, \beta_h, \underline{d}_h, \overline{d}_h) = (3, 3, 0, 7)$ and $(m_\ell, \beta_\ell, \underline{d}_\ell, \overline{d}_\ell) = (1, 10, 3, 0)$ respectively. Each buyer type θ_b, $b \in \{\ell, h\}$ picks the menu item $(m_b, \beta_b, \underline{d}_b, \overline{d}_b)$ targeted at him. After the buyer chooses, type ω_h of the seller offers a price of 10 and type ω_ℓ offers a price of 3. At any off path action chosen by the seller, the buyer assigns probability one to the seller being type ω_ℓ.

We now verify that these strategies constitute an equilibrium; in essence, they replicate the outcome from the seller-optimal equilibrium. First consider the prices offered by the seller in the last stage. We first observe that picking prices 3 and 10 results in the same outcomes as reporting ω_ℓ and ω_h respectively in the direct mechanism. To see this, observe that, if either seller type offers a price of 3, the buyer's belief assigns probability one to the seller being type ω_ℓ. The high-type buyer will choose to purchase the good since they receive no discount and their value is 3. Conversely, the low-type buyer will also purchase since he receives a discount of 3 and so the good is essentially free. Consequently, the seller's profit from offering price 3 when she is type ω_s is

$$\underbrace{\frac{1}{2} \times 3 + \frac{1}{2} \times 1}_{\text{Membership fees}} + \underbrace{\frac{1}{2} \times (3-3) + \frac{1}{2} \times 3}_{\text{Net price paid}} - c_s = 3.5 - c_s.$$

Conversely, offering price 10 yields profit

$$\underbrace{\frac{1}{2}\times 3+\frac{1}{2}\times 1}_{\text{Membership fees}}+\underbrace{\frac{1}{2}\times(10-7)}_{\text{Net price paid}}-\frac{1}{2}\times c_s=3.5-\frac{c_s}{2}.$$

In other words, offering prices 3 and 10 mimic the outcomes from the direct mechanism.

It remains to verify that neither seller type wants to deviate to any other price. Since trade happens with probability 1 at a price of 3, neither seller will deviate and offer a price of less than 3 (since this would also not affect the discounts). By offering a price t between 3 and 5, the seller's profit is

$$\underbrace{\frac{1}{2}\times 3+\frac{1}{2}\times 1}_{\text{Membership fees}}+\underbrace{\frac{1}{2}\times(t-3)+\frac{1}{2}\times(t-7)}_{\text{Net price paid}}-c_s\leq 3.5-c_s$$

and therefore this is not a profitable deviation. Offering a price t between 5 and 10 yields profit

$$\underbrace{\frac{1}{2}\times 3+\frac{1}{2}\times 1}_{\text{Membership fees}}+\underbrace{\frac{1}{2}\times(t-7)}_{\text{Net price paid}}-\frac{c_s}{2}\leq 3.5-\frac{c_s}{2}.$$

Finally, offering a price above 10 leads to a profit of

$$\underbrace{\frac{1}{2}\times 3+\frac{1}{2}\times 1}_{\text{Membership fees}}\leq 3.5-\frac{c_s}{2}.$$

Thus, each seller type ω_ℓ and ω_h offer prices 3 and 10 on path respectively. Since each buyer is infinitesimal, their decision does not affect the final price offered by the seller. By picking the menu item targeted to their type θ_b the buyer replicates the outcome from the direct mechanism; since the latter is incentive compatible, the buyer will choose their corresponding menu item. Finally, in the first stage, neither seller type ω_s has an incentive to deviate and offer any other menu since the most profit they can get from doing this is π_s. The profit from staying on path is higher since the direct mechanism satisfies the seller's deviation constraint.

5.3. *Concluding remarks*

In this paper, we study a canonical interdependent-value informed principal setting. As the title suggests, our goal is to examine the effect of seller private information on consumer surplus. We show that the buyer weakly prefers the seller to be privately informed whenever trade is inefficient if the seller has no information whatsoever. Moreover, we characterize when this preference is strict and show that typically, in this case, it is buyer-optimal for the seller information to be noisy. We view our characterization of the seller-optimal equilibrium (which is an intermediate step to deriving the buyer-optimal outcome) to be an independent contribution. We were surprised that this characterization in a canonical setting like ours was previously unknown (to the best of our knowledge).

While we view it to be natural to study the theoretical question posed in the title, as we discussed in the introduction, there is increasing scrutiny by policy makers about what information about buyers can be used by sellers to determine pricing. We view our paper to be a first step in the direction of studying this topical policy question. This issue is particularly important in the market for health insurance which is an interdependent value setting as both the insurer's value and provider's cost depend on each other's private information. Our analysis suggests that regulation needs to be nuanced as the natural urge to ban providers from using their private information can make the insurance market less efficient.

In future work, we hope to tailor our analysis to better study this specific application. One obvious generalization required to better model an insurance setting, is to introduce risk aversion. A benefit of our two-type setting is that this might be technically feasible (introducing risk-aversion frequently makes information design problems intractable). Finally, we hope to go beyond our two-type assumption although the existing informed-principal literature suggests that such a generalization will be challenging.

Appendix A. Proof of Theorem 3

First, lets us recall and explicitly write down the seller's problem, i.e. the linear program that the ex-ante mechanism must solve.

$$\max_{x,t} \sum_{b,s\in\{h,\ell\}} g_s f_b \left(t_{bs} - c_{bs}x_{bs}\right), \tag{OPT}$$

$$g_h\left(u_{hh}x_{hh} - t_{hh}\right) + g_\ell(u_{h\ell}x_{h\ell} - t_{h\ell}) \geq g_h(u_{hh}x_{\ell h} - t_{\ell h}) + g_\ell(u_{h\ell}x_{\ell\ell} - t_{\ell\ell}), \tag{B-IC-hl}$$

$$g_h\left(u_{hh}x_{hh} - t_{hh}\right) + g_\ell(u_{h\ell}x_{h\ell} - t_{h\ell}) \geq 0, \tag{B-IR-h}$$

$$g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(u_{\ell\ell}x_{\ell\ell} - t_{\ell\ell}) \geq g_h(u_{\ell h}x_{hh} - t_{hh}) + g_\ell(u_{\ell\ell}x_{h\ell} - t_{h\ell}), \tag{B-IC-lh}$$

$$g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(u_{\ell\ell}x_{\ell\ell} - t_{\ell\ell}) \geq 0, \tag{B-IR-l}$$

$$f_h(t_{h\ell} - c_{h\ell}x_{h\ell}) + f_\ell(t_{\ell\ell} - c_{\ell\ell}x_{\ell\ell}) \geq f_h(t_{hh} - c_{h\ell}x_{hh}) + f_\ell(t_{\ell h} - c_{\ell\ell}x_{\ell h}), \tag{S-IC-lh}$$

$$f_h(t_{h\ell} - c_{h\ell}x_{h\ell}) + f_\ell(t_{\ell\ell} - c_{\ell\ell}x_{\ell\ell}) \geq 0, \tag{S-IR-l}$$

$$f_h(t_{hh} - c_{hh}x_{hh}) + f_\ell(t_{\ell h} - c_{\ell h}x_{\ell h}) \geq f_h(t_{h\ell} - c_{hh}x_{h\ell}) + f_\ell(t_{\ell\ell} - c_{\ell h}x_{\ell\ell}), \tag{S-IC-hl}$$

$$f_h(t_{hh} - c_{hh}x_{hh}) + f_\ell(t_{\ell h} - c_{\ell h}x_{\ell h}) \geq 0, \tag{S-IR-h}$$

$$0 \leq x_{hh}, x_{\ell h}, x_{h\ell}, x_{\ell\ell} \leq 1.$$

Notice, that here we use the original seller-IC constraint and do no include the no deviation constraints yet. We will prove Theorem 3 in two steps. We will first characterize an ex-ante optimal mechanism, that is a solution to the linear program (OPT), to then establish that this can be supported as an equilibrium in the informed principal game.

We begin by proving some helpful (but straightforward) Lemmata. The first is the natural generalization to this multidimensional setting of the standard result that incentive compatibility requires monotonicity of the allocation rule.

Lemma 1. *In any feasible solution (x, t) to (OPT), we must have:*

$$g_h(u_{hh} - u_{\ell h})(x_{hh} - x_{\ell h}) + g_\ell(u_{h\ell} - u_{\ell\ell})(x_{h\ell} - x_{\ell\ell}) \geq 0, \tag{B-MON}$$

$$f_\ell(c_{\ell\ell} - c_{\ell h})(x_{\ell h} - x_{\ell\ell}) + f_h(c_{h\ell} - c_{hh})(x_{hh} - x_{h\ell}) \geq 0. \tag{S-MON}$$

Proof. The first inequality follows simply by adding (B-IC-hl) and (B-IC-lh) and collecting terms. Note that transfers cancel out leaving exactly the desired inequality in terms of the allocation. The second inequality follows analogously by adding (S-IC-hl) and (S-IC-lh) and collecting terms. ■

For the two-cost case discussed in Theorem 1 the second condition implies that the low (cost) type seller must have a higher expected allocation.

The next lemma establishes that in any optimal solution the high-to-low IC constraint for each player must bind (in case of the seller this is (S-IC-lh) since the low (cost) type seller is the strong type) so does the IR constraint for the low-type buyer. Moreover, the former implies that the respective low-to-high IC constraints and IR constraints for the high types for buyer and seller are satisfied.

LEMMA 2. *In any optimal solution to* (OPT), (B-IC-hl), (S-IC-lh) *and* (B-IR-l) *must bind. Moreover,* (B-IC-hl) *(respectively,* (S-IC-lh)*) binding in a feasible solution implies that* (B-IR-h) *and* (B-IC-lh) *(respectively,* (S-IR-l) *and* (S-IC-hl)*) are satisfied.*

PROOF. Suppose not, suppose (B-IC-hl) is slack in an optimal solution. Note that then (B-IR-h) must also be slack. To see this observe that:

$$\begin{aligned}
&g_h\,(u_{hh}x_{hh} - t_{hh}) + g_\ell(u_{h\ell}x_{h\ell} - t_{h\ell}) && \\
>&g_h(u_{hh}x_{\ell h} - t_{\ell h}) + g_\ell(u_{h\ell}x_{\ell\ell} - t_{\ell\ell}), && \text{(by assumption)}\\
\geq&g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(u_{\ell\ell}x_{\ell\ell} - t_{\ell\ell}), && \text{(since } u_{hh} > u_{\ell h},\ u_{h\ell} > u_{\ell\ell}\text{)}\\
\geq&0. && \text{(by (B-IR-l)).}
\end{aligned}$$

Now, consider increasing both $t_{h\ell}$ and t_{hh} by some small $\varepsilon > 0$. First, notice that this strictly increases the objective function value. Next notice that for sufficiently small ε, this continues to satisfy (B-IC-hl) and (B-IR-h). Further, this relaxes (B-IC-lh), (S-IR-l) and (S-IR-h) and leaves the other constraints unaffected. Therefore for ε small enough this perturbation is feasible and achieves a strictly higher objective function value, contradicting the optimality of the initial solution. Therefore (B-IC-hl) binds in an optimal solution.

Next, suppose (B-IR-l) is slack in an optimal solution. Consider a solution that increases all of $t_{\ell\ell}, t_{h\ell}, t_{\ell h}, t_{hh}$ by some small $\varepsilon > 0$. Observe that this strictly increases the objective function value. Further, this change leaves the buyer and seller IC constraints unaffected and relaxes the seller IR constraints. Note also that by the argument above, we can find $\varepsilon > 0$ so that this perturbation still satifies the buyer IR constraints. This contradicts optimality of the conjectured solution and hence (B-IR-l) binds in an optimal solution.

Finally, the fact that (B-IC-hl) binds, implies that

$$\begin{aligned}
&g_h\,(u_{hh}x_{hh} - t_{hh}) + g_\ell(u_{h\ell}x_{h\ell} - t_{h\ell}), && \\
=&g_h(u_{hh}x_{\ell h} - t_{\ell h}) + g_\ell(u_{h\ell}x_{\ell\ell} - t_{\ell\ell}), && \text{(since (B-IC-hl) binds)}\\
\geq&g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(u_{\ell\ell}x_{\ell\ell} - t_{\ell\ell}), && \text{(since } u_{hh} > u_{\ell h},\ u_{h\ell} > u_{\ell\ell}\text{)}\\
\geq&0. && \text{(by (B-IR-l)),}
\end{aligned}$$

that is, (B-IR-h) is satisfied.

Moreover,

$$\begin{aligned}
&g_h\,(u_{hh}x_{hh} - t_{hh}) + g_\ell(u_{h\ell}x_{h\ell} - t_{h\ell}) = g_h(u_{hh}x_{\ell h} - t_{\ell h}) + g_\ell(u_{h\ell}x_{\ell\ell} - t_{\ell\ell}),\\
\Longrightarrow\ &g_h(u_{hh}(x_{hh} - x_{\ell h})) + g_\ell(u_{h\ell}(x_{h\ell} - x_{\ell\ell})) = g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}),\\
\Longrightarrow\ &g_h(u_{\ell h}(x_{hh} - x_{\ell h})) + g_\ell(u_{\ell\ell}(x_{h\ell} - x_{\ell\ell})) \leq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}),\\
\Longrightarrow\ &g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(u_{\ell\ell}x_{\ell\ell} - t_{\ell\ell}) \geq g_h(u_{\ell h}x_{hh} - t_{hh}) + g_\ell(u_{\ell\ell}x_{h\ell} - t_{h\ell}),
\end{aligned}$$

where the third line follows from Lemma 1. This shows that (B-IC-lh) is satisfied.

Analogous arguments apply to showing that (S-IC-lh) binds and (S-IR-l) is satisfied in an optimal solution.

■

The following Observation is key to understanding buyer rents in this setting: it adapts the usual finding in single-dimensional mechanism design that rents to the high type are proportional to the allocation to the low type.

OBSERVATION 1. *In any optimal solution (x,t) to (OPT) the buyer gets positive ex-ante expected surplus if and only if $(x_{\ell\ell}, x_{\ell h}) \neq (0,0)$.*

PROOF. To see this note that from (B-IR-l) binding (Lemma 2), we have that

$$g_h t_{\ell h} + g_\ell t_{\ell\ell} = g_h u_{\ell h} x_{\ell h} + g_\ell u_{\ell\ell} x_{\ell\ell}.$$

Further, since (B-IC-hl) binds, substituting in the above, we have that

$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} x_{hh} + g_\ell t_{h\ell} u_{h\ell} - g_h(u_{hh} - u_{\ell h}) x_{\ell h} - g_\ell(u_{h\ell} - u_{\ell\ell}) x_{\ell\ell}.$$

Note that therefore the expected rents of the buyer equal

$$f_h\left(g_h(u_{hh} - u_{\ell h}) x_{\ell h} + g_\ell(u_{h\ell} - u_{\ell\ell}) x_{\ell\ell}\right).$$

The observation follows. ■

We are now ready to characterize the optimal allocation rule in an ex-ante optimal mechanism. We establish this result for the cases $\varphi_\ell > 0$ and $\varphi_h > 0$ separately.

A.1. *The case $\varphi_\ell > 0$*

THEOREM 4. *Suppose that*

$$g_\ell \varphi_\ell + g_h \varphi_h < 0,$$

so that an uninformed seller would optimally leave the buyer no surplus. Furthermore suppose that $\varphi_\ell > 0$.

The optimal allocation rule can be described as:

1. *The optimal allocation rule is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (1,1,0,1)$, if there exist associated feasible transfers in (OPT).*
2. *Otherwise the optimal allocation rule is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (1,1,0,\overline{x})$ where $\overline{x} \in [0,1]$ is the largest number such that associated feasible transfers do exist.*

PROOF OF THEOREM 4. Consider the following relaxed problem for the seller that ignores the seller's IC and IR constraints, and also (B-IR-h) and (B-IC-lh).

$$\max_{x,t} \sum_{b,s \in \{h,\ell\}} g_s f_b\left(t_{bs} - c_{bs} x_{bs}\right) \tag{R-OPT1}$$

$$\text{(B-IC-hl)}, \text{(B-IR-l)},$$

$$0 \leq x_{hh}, x_{\ell h}, x_{h\ell}, x_{\ell\ell} \leq 1.$$

Usual techniques tell us that the optimal to this relaxed problem (R-OPT1) is $x_{h\ell} = x_{\ell\ell} = x_{hh} = 1$, and $x_{\ell h} = 0$. To see this note first that both (B-IR-l) and (B-IC-hl) must bind in the optimal: if the former is slack we can increase $t_{\ell h}, t_{\ell\ell}$ while relaxing (B-IC-hl) and increase the objective function value. Similarly, for (B-IC-hl), we can increase $t_{hh}, t_{h\ell}$ without affecting (B-IR-l) and increase the objective function value.

Recall from the proof of 1, that (B-IC-hl) and (B-IR-l) binding implies

$$g_h t_{\ell h} + g_\ell t_{\ell\ell} = g_h u_{\ell h} x_{\ell h} + g_\ell u_{\ell\ell} x_{\ell\ell}, \quad \text{and}$$
$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} x_{hh} + g_\ell u_{h\ell} x_{h\ell} - g_h(u_{hh} - u_{\ell h})x_{\ell h} - g_\ell(u_{h\ell} - u_{\ell\ell})x_{\ell\ell}.$$

Substituting this into the objective function and collecting terms yields:

$$g_h f_h(u_{hh} - c_{hh})x_{hh} + g_h \varphi_h x_{\ell h} + g_\ell f_h(u_{h\ell} - c_{h\ell})x_{h\ell} + g_\ell f_\ell \varphi_\ell x_{\ell\ell}.$$

By the assumption that there are always gains from trade, $u_{hh} - c_{hh}, u_{h\ell} - c_{h\ell} \geq 0$, and by assumption we have that $\varphi_\ell > 0$ while $\varphi_h < 0$. Observe that in this case the pointwise optimal is $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$. If this solution is feasible in the original program (OPT), that is, if there exist transfers that satisfy *both* the buyer's *and* the seller's IC and IR constraints, then we are done.

So suppose not. For this case, we first prove the following claim.

CLAIM 1. *The pointwise optimal solution $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$ is not feasible in (OPT) if and only if $g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} < 0$.*

PROOF OF CLAIM. First, note that the pointwise optimal solution $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$ is not feasible in (OPT) if and only if there exists no solution (i.e., no transfers $t_{hh}, t_{\ell h}, t_{h\ell}, t_{\ell\ell}$) to the system of inequalities (B-IC-hl), (B-IR-l), (S-IC-lh), (S-IR-h) with $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$. To see this, note that by Lemma 2 if there exists a solution to these system of inequalities then there also exists a solution which satisfies (B-IR-h), (B-IC-lh),(S-IR-l) and (S-IC-hl).

Plugging $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$ into (B-IC-hl), (B-IR-l), (S-IC-lh), (S-IR-h) and collecting terms we have the following system. The tags of the inequalities represent the corresponding variable in the Farkas' alternative.

$$g_h u_{hh} \geq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A}$$
$$g_\ell u_{\ell\ell} \geq g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B}$$
$$f_h(t_{h\ell} - t_{hh}) + f_\ell(t_{\ell\ell} - t_{\ell h}) \geq f_\ell c_{\ell\ell}, \tag{C}$$
$$f_h t_{hh} + f_\ell t_{\ell h} \geq f_h c_{hh}. \tag{D}$$

By the Farkas Lemma, either there exists a solution to the system above or to the Farkas' alternative below, but not both:

$$g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h c_{hh} D < 0,$$
$$g_h A + f_h C - f_h D = 0, \tag{t_{hh}}$$
$$-g_h A + g_h B + f_\ell C - f_\ell D = 0, \tag{$t_{\ell h}$}$$
$$g_\ell A - f_h C = 0, \tag{$t_{h\ell}$}$$
$$-g_\ell A + g_\ell B - f_\ell C = 0, \tag{$t_{\ell\ell}$}$$
$$A, B, C, D \geq 0.$$

Here the tags of the equations represent the variable in the original system to which this equation corresponds to in the Farkas' alternative. Observe that ($t_{h\ell}$) implies that:

$$C = \frac{g_\ell}{f_h} A.$$

Plugging into ($t_{\ell\ell}$),

$$g_\ell B = g_\ell A + f_\ell \frac{g_\ell}{f_h} A,$$
$$\implies B = \frac{1}{f_h} A.$$

Plugging into (t_{hh}),

$$f_h D = g_h A + g_\ell A$$
$$\implies D = \frac{1}{f_h} A.$$

It is easy to verify that this satisfies ($t_{\ell h}$). Finally note that that since

$$g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h c_{hh} D < 0,$$
$$\implies g_h u_{hh} A + \frac{g_\ell}{f_h} u_{\ell\ell} A - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} A - c_{hh} A < 0,$$
$$\implies g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} < 0 \qquad \text{(since } A \geq 0.\text{)}$$

Thus, the Farkas' alternative is feasible if and only if $g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} < 0$. The claim follows. ■

CLAIM 2. *If $g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} < 0$, the optimal solution is $x_{\ell\ell} = x_{h\ell} = 1, x_{\ell h} = 0$ and $x_{hh} \in (0,1)$ as large as possible such that there exist transfers which satisfy* (B-IC-hl), (B-IR-l), (S-IC-lh), (S-IR-h).

PROOF. We will prove this claim by considering the relaxed problem

$$\max_{x,t} \sum_{b,s \in \{h,\ell\}} g_s f_b \left(t_{bs} - c_{bs} x_{bs}\right) \qquad \text{(R-OPT2)}$$
$$\text{s.t. (B-IC-hl), (B-IR-l), (S-IC-lh), (S-IR-h),}$$
$$0 \leq x_{hh}, x_{\ell h}, x_{h\ell}, x_{\ell\ell} \leq 1.$$

and constructing a dual solution which complements it.

We will characterize the solution to this LP in the usual way, i.e. construct the dual and show the existence of a dual feasible solution that complements the candidate primal optimal solution. To that end, let us denote that dual variables corresponding to (B-IC-hl) by β_{IC}, (B-IR-l) by β_{IR}, (S-IC-lh) by σ_{IC} and (S-IR-h) by σ_{IR}. The dual variables corresponding to the upper bounds on x are denoted by η, e.g. the dual variable corresponding to $x_{hh} \leq 1$ is denoted η_{hh}. Since all the primal constraints are inequalities, the dual variables must be non-negative.

In the dual that follows, each dual constraint is tagged with the primal variable to which it corresponds. Since the t's are unsigned in the primal, the corresponding dual constraints are equalities. Since the x's are non-negative, the corresponding dual variables are inequalities.

$$\min_{\beta,\sigma,\eta} \sum_{b,s\in\{h,\ell\}} \eta_{bs} \tag{D-OPT}$$

$$g_\ell u_{h\ell}\beta_{\text{IC}} - g_\ell u_{\ell\ell}\beta_{\text{IR}} + f_\ell c_{\ell\ell}\sigma_{\text{IC}} + \eta_{\ell\ell} \geq -g_\ell f_\ell c_{\ell\ell}, \tag{$x_{\ell\ell}$}$$

$$- g_\ell u_{h\ell}\beta_{\text{IC}} + f_h c_{h\ell}\sigma_{\text{IC}} + \eta_{h\ell} \geq -g_\ell f_h c_{h\ell}, \tag{$x_{h\ell}$}$$

$$g_h u_{hh}\beta_{\text{IC}} - g_h u_{\ell h}\beta_{\text{IR}} - f_\ell c_{\ell\ell}\sigma_{\text{IC}} + f_\ell c_{\ell h}\sigma_{\text{IR}} + \eta_{\ell h} \geq -g_h f_\ell c_{\ell h}, \tag{$x_{\ell h}$}$$

$$- g_h u_{hh}\beta_{\text{IC}} - f_h c_{h\ell}\sigma_{\text{IC}} + f_h c_{hh}\sigma_{\text{IR}} + \eta_{hh} \geq -g_h f_h c_{hh}, \tag{x_{hh}}$$

$$- g_\ell\beta_{\text{IC}} + g_\ell\beta_{\text{IR}} - f_\ell\sigma_{\text{IC}} = g_\ell f_\ell, \tag{$t_{\ell\ell}$}$$

$$g_\ell\beta_{\text{IC}} - f_h\sigma_{\text{IC}} = g_\ell f_h, \tag{$t_{h\ell}$}$$

$$- g_h\beta_{\text{IC}} + g_h\beta_{\text{IR}} + f_\ell\sigma_{\text{IC}} - f_\ell\sigma_{\text{IR}} = g_h f_\ell, \tag{$t_{\ell h}$}$$

$$g_h\beta_{\text{IC}} + f_h\sigma_{\text{IC}} - f_h\sigma_{\text{IR}} = g_h f_h, \tag{t_{hh}}$$

$$\beta_{\text{IC}}, \beta_{\text{IR}}, \sigma_{\text{IC}}, \sigma_{\text{IR}}, \eta_{\ell\ell}, \eta_{h\ell}, \eta_{\ell h}, \eta_{hh} \geq 0.$$

From the dual constraints corresponding to $(t_{\ell\ell}, t_{h\ell}, t_{\ell h}, t_{hh})$ we can now achieve some simplification. First, from ($t_{h\ell}$), we have that:

$$\sigma_{\text{IC}} = \frac{g_\ell}{f_h}\beta_{\text{IC}} - g_\ell. \tag{4}$$

Substituting into ($t_{\ell\ell}$) we have that

$$\beta_{\text{IR}} = \frac{\beta_{\text{IC}}}{f_h}. \tag{5}$$

Substituting (4) into (t_{hh}), we have that

$$\sigma_{\text{IR}} = \frac{\beta_{\text{IC}}}{f_h} - 1. \tag{6}$$

Observe that dual feasibility therefore requires that $\frac{\beta_{\text{IC}}}{f_h} - 1 \geq 0$. We can now substitute (4–6) into the dual constraints ($x_{\ell\ell}$,$x_{h\ell}$,$x_{\ell h}$,x_{hh}), yielding:

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{h\ell}\frac{f_h}{f_\ell} - u_{\ell\ell}\frac{1}{f_\ell} + c_{\ell\ell}\right) + \frac{\eta_{\ell\ell}}{f_\ell g_\ell} \geq 0, \tag{$x_{\ell\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-u_{h\ell} + c_{h\ell}\right) + \frac{\eta_{h\ell}}{f_h g_\ell} \geq 0, \tag{$x_{h\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{hh}\frac{f_h}{f_\ell} - u_{\ell h}\frac{1}{f_\ell} + c_{\ell h} + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} + \frac{\eta_{\ell h}}{f_\ell g_h} \geq 0, \tag{$x_{\ell h}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right) - (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h} + \frac{\eta_{hh}}{f_h g_h} \geq 0. \tag{x_{hh}'}$$

Consider the following solution:

$$\frac{\beta_{\text{IC}}}{f_h} = \frac{(c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}}{\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right)} = \frac{(c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}}{\left(-u_{hh} + c_{hh}\frac{1}{g_h} - c_{h\ell}\frac{g_\ell}{g_h}\right)} \tag{7}$$

Recall that we consider the case $\varphi_\ell = u_{\ell\ell} - u_{h\ell} f_h - c_{\ell\ell} f_\ell > 0$ and $g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} < 0$ (Claim 2). This implies

$$\implies g_h u_{hh} + g_\ell u_{h\ell} - c_{hh} < 0 \qquad \text{(since } \varphi_\ell \geq 0\text{)}$$
$$\implies g_h u_{hh} + g_\ell c_{h\ell} - c_{hh} < 0 \qquad \text{(since } u_{h\ell} \geq c_{h\ell}\text{)}$$

Consequently, the denominator on the right hand side of (7) is positive, $c_{hh} - c_{h\ell} \geq 0$ by assumption and therefore $\frac{\beta_{\text{IC}}}{f_h} \geq 0$. Further, since $u_{hh} - c_{hh} \geq 0$ by assumption, we further have that $\frac{\beta_{\text{IC}}}{f_h} \geq 1$ and thus $\beta_{\text{IR}}, \sigma_{\text{IC}}, \sigma_{\text{IR}} \geq 0$, so this is a feasible (partial) solution. Further, we can set:

$$\eta_{h\ell} = \frac{\beta_{\text{IC}}}{f_h}(u_{h\ell} - c_{h\ell}) f_h g_\ell \geq 0,$$
$$\eta_{\ell\ell} = \frac{\beta_{\text{IC}}}{f_h f_\ell} \varphi_\ell f_\ell g_\ell \geq 0,$$
$$\eta_{hh} = \eta_{\ell h} = 0.$$

Observe that this proposed solution is therefore dual feasible and complements the proposed primal solution. To see the latter, observe that the dual constraints corresponding to $x_{h\ell}, x_{\ell\ell}$ and x_{hh} bind by construction.

Further, observe that the left hand side of the dual constraint corresponding to $x_{\ell h}$

$$\begin{aligned}
&= \frac{\beta_{\text{IC}}}{f_h}\left(u_{hh}\frac{f_h}{f_\ell} - u_{\ell h}\frac{1}{f_\ell} + c_{\ell h} + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} \\
&= \frac{\beta_{\text{IC}}}{f_h f_\ell}(-\varphi_h + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell f_\ell}{g_h}) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} \\
&\geq \frac{\beta_{\text{IC}}}{f_h f_\ell}(-\varphi_h) && \text{(since } \frac{\beta_{\text{IC}}}{f_h} \geq 1, c_{\ell h} \geq c_{\ell\ell}\text{)} \\
&\geq 0 && \text{(since } \varphi_h < 0 \text{ by condition A)}
\end{aligned}$$

Therefore we have a dual-feasible solution that complements our proposed primal solution, witnessing its optimality. ■

This also completes the proof of Theorem 4. ■

A.2. *The case $\varphi_\ell < 0$*

THEOREM 5. *Suppose that*

$$g_\ell \varphi_\ell + g_h \varphi_h < 0,$$

so that an uninformed seller would optimally leave the buyer no surplus. Furthermore suppose we have that $\varphi_h > 0$*. The optimal allocation rule can be described as follows:*

(1) *Case 1:* $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h \leq f_h(u_{hh}-c_{hh})$.
   (a) *If there exist transfers that are feasible in* (OPT)*, the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,0,1)$.
   (b) *Otherwise, the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,0,\bar{x})$*, where* $\bar{x}$ *is the largest number in* $[0,1]$ *such that a solution does exist.*

(2) *Case 2:* $\frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h > f_h(u_{hh}-c_{hh})$*. Define* $k := \frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}$.
   (a) $k < 1$*:*
      (i) *If there exist transfers that are feasible in* (OPT)*, the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,x_{\ell h},0)$*, where* $x_{\ell h}$ *solves:*

$$f_\ell(c_{\ell\ell}-c_{\ell h})(x_{\ell h}) + f_h(c_{h\ell}-c_{hh})(-1) = 0.$$

      (ii) *Otherwise the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,\bar{x},0)$*, where* $\bar{x}$ *is the largest number in* $[0,1]$ *such that a solution does exist.*
   (b) $k > 1$*:*
      (i) *If there exist transfers that are feasible in* (OPT)*, the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,1,x_{hh})$*, where* y *solves*

$$f_\ell(c_{\ell\ell}-c_{\ell h}) + f_h(c_{h\ell}-c_{hh})(x_{hh}-1) = 0.$$

      (ii) *Otherwise, the optimal allocation rule is:* $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0,1,1,\bar{x})$*, where* $\bar{x}$ *is the largest number in* $[0,1]$ *such that a solution does exist.*

PROOF. We will again show the characterize the optimal allocation/solution by constructing a dual solution to (OPT) which complements it. By Lemma 2, we have that (B-IC-hl) and (S-IC-lh) bind in any optimal solution. By observation we can also have that (S-IC-hl) binds (since a candidate optimal allocation could be that $(x_{\ell h}, x_{hh}) = (x_{\ell\ell}, x_{h\ell}) = (0,1)$ in the proposed solution).[6] Consider therefore the relaxed primal program:

$$\max_{x,t} \sum_{b,s\in\{h,\ell\}} g_s f_b\left(t_{bs} - c_{bs}x_{bs}\right) \tag{R-OPT3}$$

$$\text{s.t. (B-IC-hl), (B-IR-l), (S-IC-lh), (S-IC-hl), (S-IR-h),}$$
$$0 \leq x_{hh}, x_{\ell h}, x_{h\ell}, x_{\ell\ell} \leq 1.$$

By observation (cf. 2), if we can find transfers that are feasible here, they will also be feasible in the initial program (OPT).

To that end, let us denote the dual variables corresponding to (B-IC-hl) by β_{IC}, (B-IR-l) by β_{IR}, (S-IC-lh) by σ_{IC}, (S-IC-hl) by σ'_{IC} and (S-IR-h) by σ_{IR}. As previously, the dual variables corresponding to the upperbounds on x are denoted by η, e.g. the dual variable corresponding to $x_{hh} \leq 1$ is denoted η_{hh}. Since all the primal constraints are inequalities, the dual variables must be nonnegative.

[6] Notice that this is also the optimal allocation when the seller has no information about q.

In the dual that follows, each dual constraint is tagged with the primal variable to which it corresponds. Since the t's are unsigned in the primal, the corresponding dual constraints are equalities. Since the x's are non-negative, the corresponding dual variables are inequalities.

$$\min \sum_{b,s\in\{h,\ell\}} \eta_{bs} \tag{D-OPT2}$$

$$g_\ell u_{h\ell}\beta_{\text{IC}} - g_\ell u_{\ell\ell}\beta_{\text{IR}} + f_\ell c_{\ell\ell}\sigma_{\text{IC}} - f_\ell c_{\ell h}\sigma'_{\text{IC}} + \eta_{\ell\ell} \geq -g_\ell f_\ell c_{\ell\ell}, \tag{$x_{\ell\ell}$}$$

$$-g_\ell u_{h\ell}\beta_{\text{IC}} + f_h c_{h\ell}\sigma_{\text{IC}} - f_h c_{hh}\sigma'_{\text{IC}} + \eta_{h\ell} \geq -g_\ell f_h c_{h\ell}, \tag{$x_{h\ell}$}$$

$$g_h u_{hh}\beta_{\text{IC}} - g_h u_{\ell h}\beta_{\text{IR}} - f_\ell c_{\ell\ell}\sigma_{\text{IC}} + f_\ell c_{\ell h}\sigma'_{\text{IC}} + f_\ell c_{\ell h}\sigma_{\text{IR}} + \eta_{\ell h} \geq -g_h f_\ell c_{\ell h}, \tag{$x_{\ell h}$}$$

$$-g_h u_{hh}\beta_{\text{IC}} - f_h c_{h\ell}\sigma_{\text{IC}} + f_h c_{hh}\sigma'_{\text{IC}} + f_h c_{hh}\sigma_{\text{IR}} + \eta_{hh} \geq -g_h f_h c_{hh}, \tag{x_{hh}}$$

$$-g_\ell\beta_{\text{IC}} + g_\ell\beta_{\text{IR}} - f_\ell\sigma_{\text{IC}} + f_\ell\sigma'_{\text{IC}} = g_\ell f_\ell, \tag{$t_{\ell\ell}$}$$

$$g_\ell\beta_{\text{IC}} - f_h\sigma_{\text{IC}} + f_h\sigma'_{\text{IC}} = g_\ell f_h, \tag{$t_{h\ell}$}$$

$$-g_h\beta_{\text{IC}} + g_h\beta_{\text{IR}} + f_\ell\sigma_{\text{IC}} - f_\ell\sigma'_{\text{IC}} - f_\ell\sigma_{\text{IR}} = g_h f_\ell, \tag{$t_{\ell h}$}$$

$$g_h\beta_{\text{IC}} + f_h\sigma_{\text{IC}} - f_h\sigma'_{\text{IC}} - f_h\sigma_{\text{IR}} = g_h f_h, \tag{t_{hh}}$$

$$\beta_{\text{IC}}, \beta_{\text{IR}}, \sigma_{\text{IC}}, \sigma'_{\text{IC}}, \sigma_{\text{IR}}, \eta_{\ell\ell}, \eta_{h\ell}, \eta_{\ell h}, \eta_{hh} \geq 0.$$

From the dual constraints corresponding to $(t_{\ell\ell}, t_{h\ell}, t_{\ell h}, t_{hh})$ we can now achieve some simplification. First, from ($t_{h\ell}$), we have that:

$$\sigma_{\text{IC}} - \sigma'_{\text{IC}} = g_\ell\left(\frac{\beta_{\text{IC}}}{f_h} - 1\right). \tag{8}$$

Substituting into ($t_{\ell\ell}$) we have that

$$\beta_{\text{IR}} = \frac{\beta_{\text{IC}}}{f_h}. \tag{9}$$

Substituting (8) into (t_{hh}), we have that

$$\sigma_{\text{IR}} = \frac{\beta_{\text{IC}}}{f_h} - 1 \tag{10}$$

We can now substitute (8–10) into the dual constraints ($x_{\ell\ell}$,$x_{h\ell}$,$x_{\ell h}$,x_{hh}), yielding:

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{h\ell}\frac{f_h}{f_\ell} - u_{\ell\ell}\frac{1}{f_\ell} + c_{\ell\ell}\right) - \frac{1}{g_\ell}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} + \frac{\eta_{\ell\ell}}{g_\ell f_\ell} \geq 0, \tag{$x_{\ell\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}(-u_{h\ell} + c_{h\ell}) - \frac{1}{g_\ell}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} + \frac{\eta_{h\ell}}{g_\ell f_h} \geq 0, \tag{$x_{h\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{hh}\frac{f_h}{f_\ell} - u_{\ell h}\frac{1}{f_\ell} + c_{\ell h} + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) + \frac{1}{g_h}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} + \frac{\eta_{\ell h}}{g_h f_\ell} \geq 0, \tag{$x_{\ell h}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right) + \frac{1}{g_h}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} - (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h} + \frac{\eta_{hh}}{g_h f_h} \geq 0, \tag{x_{hh}'}$$

Case 1.

Case 1a: $\frac{1}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h \leq \frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})}$ and $g_h u_{hh} + g_\ell u_{h\ell} \geq c_{hh}$.

We will show that in this case the optimal solutions is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, 0, 1)$.

Observe that the second condition implies that the high-cost seller would be willing to sell at a price being the expected value of the high-type buyer. Thus (S-IR-h) is slack and hence $\sigma_{\text{IR}} = 0$; (10) then implies $\frac{\beta_{\text{IC}}}{f_h} = 1$. Substituting in and applying complementary slackness we need to show that we can construct a solution to

$$
\begin{aligned}
& -\frac{1}{f_\ell}\varphi_\ell - \frac{1}{g_\ell}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, && (x_{\ell\ell}{}') \\
& (-u_{h\ell} + c_{h\ell}) - \frac{1}{g_\ell}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} + \frac{\eta_{h\ell}}{g_\ell f_h} = 0, && (x_{h\ell}{}') \\
& -\frac{1}{f_\ell}\varphi_h + \frac{1}{g_h}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, && (x_{\ell h}{}') \\
& (-u_{hh} + c_{hh}) + \frac{1}{g_h}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} + \frac{\eta_{hh}}{g_h f_h} = 0, && (x_{hh}{}') \\
& \sigma'_{\text{IC}}, \eta_{hh}, \eta_{h\ell} \geq 0
\end{aligned}
$$

Note that the second equality $(x'_{h\ell})$ must have a solution because the first two terms are non-positive. So we are left to show there is a solution to:

$$
\begin{aligned}
& -\frac{1}{f_\ell}\varphi_\ell - \frac{1}{g_\ell}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, && (x_{\ell\ell}{}') \\
& -\frac{1}{f_\ell}\varphi_h + \frac{1}{g_h}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, && (x_{\ell h}{}') \\
& (-u_{hh} + c_{hh}) + \frac{1}{g_h}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} \leq 0, && (x_{hh}{}') \\
& \sigma'_{\text{IC}} \geq 0
\end{aligned}
$$

Observe that $\sigma'_{\text{IC}} = \frac{g_h}{f_\ell}\frac{\varphi_h}{(c_{\ell h} - c_{\ell\ell})}$ is positive by our assumptions and solves all three inequalities (where the last one follows from $\frac{1}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h \leq \frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})}$).

Case 1b: $\frac{1}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h \leq \frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})}$ and $g_h u_{hh} + g_\ell u_{h\ell} < c_{hh}$.

We will show that in this case the optimal solution is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, 0, x)$ where x is the largest $x \in [0, 1]$ such that (S-IR-h) binds.

Plugging in the corresponding complementary slackness into our dual constraints we need to show that we can construct a solution to

$$
\begin{aligned}
& -\frac{\beta_{\text{IC}}}{f_h f_\ell}\varphi_\ell \geq 0, && (x_{\ell\ell}{}') \\
& \frac{\beta_{\text{IC}}}{f_h}(-u_{h\ell} + c_{h\ell}) \leq 0, && (x_{h\ell}{}') \\
& \frac{\beta_{\text{IC}}}{f_h}\left(-\frac{1}{f_\ell}\varphi_h + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} \geq 0, && (x_{\ell h}{}') \\
& \frac{\beta_{\text{IC}}}{f_h}\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right) - (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h} = 0, && (x_{hh}{}') \\
& \beta_{\text{IC}} \geq 0.
\end{aligned}
$$

Notice that for the first inequality we use ($x'_{h\ell}$) binding (complementary slackness) and $\eta_{h\ell} \geq 0$. Note that the inequalities ($x'_{\ell\ell}$) and ($x'_{h\ell}$) are satisfied for any nonnegative value of β_{IC}, while (x'_{hh}) can be satisfied since we assumed $c_{hh} > g_h u_{hh} + g_\ell u_{h\ell} \geq g_h u_{hh} + g_\ell c_{h\ell}$. Notice that this condition also implies $-\frac{1}{f_\ell}\varphi_h + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}$ and hence ($x'_{\ell h}$) can be satisfied. Indeed,

$$-\frac{1}{f_\ell}\varphi_h + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} \geq \left[-\frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})} + \frac{g_\ell}{g_h}\right](c_{\ell h} - c_{\ell\ell}) \geq 0$$

where the last equality follows from $g_h u_{hh} + g_\ell c_{h\ell} < c_{hh}$.

Case 2. Note that by assumption in this case we have that $\frac{1}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h > \frac{(u_{hh}-c_{hh})}{(c_{hh}-c_{h\ell})}$

In this case, if it is feasible, the solution is either of the form $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x, 0)$ or of the form $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, 1, y)$ depending on whether $k := \frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})}$ is larger than or smaller than 1.

Case 2a: Consider the case with $\frac{1}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h > \frac{(u_{hh}-c_{hh})}{(c_{hh}-c_{h\ell})}$ and $k = \frac{f_h(c_{hh}-c_{h\ell})}{f_\ell(c_{\ell h}-c_{\ell\ell})} < 1$.

We will show that in this case the optimal allocation is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x_{\ell h}, 0)$ where $x_{\ell h} \in [0, 1]$ is determined by either (S-IC-hl) (or in other words the seller's monotonicity constraint) binding, (Case 2a (i)) or by (S-IR-h) binding (Case 2a (ii)).

Notice that for the allocation $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x_{\ell h}, 0)$, the seller's monotonicity constraint (S-MON) is binding iff $x_{\ell h}$ solves

$$f_\ell(c_{\ell\ell} - c_{\ell h})(x_{\ell h}) + f_h(c_{h\ell} - c_{hh})(-1) = 0. \tag{11}$$

Moreover, observe that in an optimal solutions with allocation $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x_{\ell h}, 0)$ either (S-IR-h) or (S-IC-hl) can be binding but not both. Indeed, if (S-IR-h) is binding, then (S-IC-lh) will be slack since the low-cost seller must obtain a positive profit by assumption (there are gains from trade).

Case 2a (i): Let us consider the case in which there exists an solution with $x_{\ell h} \in [0, 1]$ that satisfies (11) for which (S-IR-h) doesn't bind. In this case, $\sigma_{\text{IR}} = 0$ in the dual solution by complementary slackness. So we must have from (10) that $\frac{\beta_{\text{IC}}}{f_h} = 1$. By complementary slackness, since our candidate solution is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x, 0)$ we must have that $\eta_{\ell\ell}, \eta_{\ell h}, \eta_{hh} = 0$ and that the dual equations corresponding to $x_{h\ell}, x_{\ell h}$ must bind. Substituting into the dual constraints we need to show that there exists a solution to:

$$\left(-\varphi_\ell \frac{1}{f_\ell}\right) - \frac{1}{g_\ell}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, \tag{$x_{\ell\ell}'$}$$

$$(-u_{h\ell} + c_{h\ell}) - \frac{1}{g_\ell}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} \leq 0, \tag{$x_{h\ell}'$}$$

$$\left(-\varphi_h \frac{1}{f_\ell}\right) + \frac{1}{g_h}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} = 0, \tag{$x_{\ell h}'$}$$

$$(-u_{hh} + c_{hh}) + \frac{1}{g_h}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} \geq 0 \tag{x_{hh}'}$$

$$\sigma'_{\text{IC}} \geq 0.$$

Observe that plugging from ($x'_{\ell h}$) we obtain

$$\sigma'_{\text{IC}} = \frac{g_h \varphi_h}{f_\ell(c_{\ell h} - c_{\ell\ell})}.$$

This is positive and therefore feasible since $\varphi_h > 0$ by assumption. We are left to show that the following inequalities are satisifed:

$$-\frac{\varphi_\ell}{f_\ell} - \frac{g_h}{g_\ell f_\ell}\varphi_h \geq 0, \tag{$x_{\ell\ell}'$}$$

$$(-u_{h\ell} + c_{h\ell}) - \frac{g_h}{g_\ell}\frac{(c_{hh} - c_{h\ell})}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h \leq 0, \tag{$x_{h\ell}'$}$$

$$(-u_{hh} + c_{hh}) + \frac{(c_{hh} - c_{h\ell})}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h \geq 0. \tag{x_{hh}'}$$

Now, the first of these ($x'_{\ell\ell}$) follows since $g_h\varphi_h + g_\ell\varphi_\ell < 0$ by assumption. Further the last of these (x'_{hh}) is satisfied since we consider the case $\frac{1}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h > \frac{(u_{hh}-c_{hh})}{(c_{hh}-c_{h\ell})}$. Finally note that ($x'_{h\ell}$) is trivially satisfied since both terms are negative.

Case 2a(ii). Suppose now that (S-IR-h) binds. In this case, $\sigma_{\text{IR}} \geq 0$ in the dual solution by complementary slackness. Conversely, seller monotonicity does not bind by assumption so $\sigma'_{\text{IC}} = 0$. By complementary slackness, since our candidate solution is $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x, 0)$ we must have that $\eta_{\ell\ell}, \eta_{\ell h}, \eta_{hh} = 0$ and that the dual equations corresponding to $x_{h\ell}, x_{\ell h}$ must bind in the dual. Substituting into the dual constraints we need to show that there exists a solution to:

$$-\frac{\beta_{\text{IC}}}{f_h f_\ell}\varphi_\ell \geq 0, \tag{$x_{\ell\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}(-u_{h\ell} + c_{h\ell}) \leq 0, \tag{$x_{h\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-\frac{1}{f_\ell}\varphi_h + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} = 0, \tag{$x_{\ell h}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right) - (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h} \geq 0 \tag{x_{hh}'}$$

$$\sigma_{\text{IR}}, \beta_{\text{IC}} \geq 0.$$

Observe that by (10) $\sigma_{\text{IR}} \geq 0$ requires that the solution of the dual satisfies $\frac{\beta_{\text{IC}}}{f_h} \geq 1$. Notice moreover that the two inequalities $x'_{\ell\ell}$ and $x'_{h\ell}$ are satisfied for any $\beta_{\text{IC}} \geq 0$. Rewriting the latter two dual constraints, we obtain:

$$\frac{\beta_{\text{IC}}}{f_h}\left(1 - \frac{\frac{1}{f_\ell}\varphi_h}{(c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}}\right) = 1., \tag{$x_{\ell h}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(1 - \frac{u_{hh} - c_{hh}}{(c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}}\right) \geq 1, \tag{x_{hh}'}$$

The solution equality $x'_{\ell h}$ satisfies x'_{hh} since $\frac{1}{f_\ell(c_{\ell h}-c_{\ell\ell})}\varphi_h > \frac{(u_{hh}-c_{hh})}{(c_{hh}-c_{h\ell})}$. We are left to verify that the solution indeed satisfies $\beta_{\text{IC}}/f_h \geq 1$, which here is equivalent to requiring that $\frac{\frac{1}{f_\ell}\varphi_h}{(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}} \leq 1$. We will now verify the latter.

Recall that, in an optimal solutions with allocation $(x_{\ell\ell}, x_{h\ell}, x_{\ell h}, x_{hh}) = (0, 1, x_{\ell h}, 0)$ either (S-IR-h) or (S-IC-hl) are binding but not both. Since we are considering the case where (S-IR-h) binds, this observation implies that there is no solution to (here we plugged the allocation into the constraints):

$$g_h(-t_{hh}) + g_\ell(u_{h\ell} - t_{h\ell}) \geq g_h(u_{hh}x_{\ell h} - t_{\ell h}) + g_\ell(-t_{\ell\ell}), \tag{B-IC-hl}$$

$$g_h(u_{\ell h}x_{\ell h} - t_{\ell h}) + g_\ell(-t_{\ell\ell}) \geq 0, \tag{B-IR-l}$$

$$f_h(t_{h\ell} - c_{h\ell}) + f_\ell(t_{\ell\ell}) \geq f_h(t_{hh}) + f_\ell(t_{\ell h} - c_{\ell\ell}x_{\ell h}), \tag{S-IC-lh}$$

$$f_h(t_{hh}) + f_\ell(t_{\ell h} - c_{\ell h}x_{\ell h}) \geq 0, \tag{S-IR-h}$$

(12)

with $x_{\ell h} = \frac{f_h(c_{h\ell}-c_{hh})}{f_\ell(c_{\ell\ell}-c_{\ell h})}$, i.e. with the seller's monotonicity constraint (S-MON) binding.

We will show that there is no solutions to this system if $\frac{\frac{1}{f_\ell}\varphi_h}{(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}} \leq 1$ and hence $\beta_{\text{IC}}/f_h \geq 1$.

We can rewrite this as:

$$-g_h u_{hh} x_{\ell h} + g_\ell u_{h\ell} \geq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A}$$

$$g_h u_{\ell h} x_{\ell h} \geq g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B}$$

$$-f_h c_{h\ell} + f_\ell c_{\ell\ell} x_{\ell h} \geq f_h(t_{hh} - t_{h\ell}) + f_\ell(t_{\ell h} - t_{\ell\ell}), \tag{C}$$

$$-f_\ell c_{\ell h} x_{\ell h} \geq -f_h t_{hh} - f_\ell t_{\ell h}. \tag{D}$$

By the Farkas Lemma, either there exists a solution to the system above or to the Farkas alternative below, but not to both:

$$(-g_h u_{hh} x_{\ell h} + g_\ell u_{h\ell})A + g_h u_{\ell h} x_{\ell h} B - (f_h c_{h\ell} - f_\ell c_{\ell\ell} x_{\ell h})C - (f_\ell c_{\ell h} x_{\ell h})D < 0,$$

$$g_h A + f_h C - f_h D = 0, \tag{t_{hh}}$$

$$-g_h A + g_h B + f_\ell C - f_\ell D = 0, \tag{$t_{\ell h}$}$$

$$g_\ell A - f_h C = 0, \tag{$t_{h\ell}$}$$

$$-g_\ell A + g_\ell B - f_\ell C = 0, \tag{$t_{\ell\ell}$}$$

$$A, B, C, D \geq 0.$$

Observe that ($t_{h\ell}$) implies that:

$$C = \frac{g_\ell}{f_h}A.$$

Plugging into ($t_{\ell\ell}$),

$$g_\ell B = g_\ell A + f_\ell \frac{g_\ell}{f_h} A,$$

$$\implies B = \frac{1}{f_h} A.$$

Plugging into (t_{hh}),

$$f_h D = g_h A + g_\ell A$$
$$\implies D = \frac{1}{f_h} A.$$

It is easy to verify that this satisfies ($t_{\ell h}$).

Finally note that

$$\begin{aligned}
0 >& (-g_h u_{hh} x_{\ell h} + g_\ell u_{h\ell})A + g_h u_{\ell h} x_{\ell h} B - (f_h c_{h\ell} - f_\ell c_{\ell\ell} x_{\ell h})C - (f_\ell c_{\ell h} x_{\ell h})D \\
=& (-g_h u_{hh} x_{\ell h} + g_\ell u_{h\ell})A + g_h u_{\ell h} x_{\ell h} \frac{A}{f_h} - (f_h c_{h\ell} - f_\ell c_{\ell\ell} x_{\ell h}) \frac{g_\ell A}{f_h} - (f_\ell c_{\ell h} x_{\ell h}) \frac{A}{f_h} \\
=& \left(-g_h u_{hh} x_{\ell h} + g_\ell u_{h\ell} + \frac{g_h}{f_h} u_{\ell h} x_{\ell h} - g_\ell c_{h\ell} + \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} x_{\ell h} - \frac{f_\ell}{f_h} c_{\ell h} x_{\ell h} \right) A \\
=& \left(\frac{g_h}{f_h} \varphi_h x_{\ell h} + g_\ell (u_{h\ell} - c_{h\ell}) + \frac{f_\ell g_\ell}{f_h} (c_{\ell\ell} - c_{\ell h}) x_{\ell h} \right) A \\
=& \left(\left(\frac{g_h}{f_h} \varphi_h - \frac{f_\ell g_\ell}{f_h} (c_{\ell h} - c_{\ell\ell}) \right) x_{\ell h} + g_\ell (u_{h\ell} - c_{h\ell}) \right) A
\end{aligned}$$

Since $A \geq 0$, the first factor must be negative. Plugging in that $x_{\ell h} = \frac{f_h(c_{h\ell} - c_{hh})}{f_\ell(c_{\ell\ell} - c_{\ell h})}$, we obtain

$$\begin{aligned}
& \left(\frac{g_h}{f_h} \varphi_h - \frac{f_\ell g_\ell}{f_h} (c_{\ell h} - c_{\ell\ell}) \right) \frac{f_h(c_{h\ell} - c_{hh})}{f_\ell(c_{\ell\ell} - c_{\ell h})} + g_\ell (u_{h\ell} - c_{h\ell}) < 0 \\
\implies & \left(g_h \varphi_h - f_\ell g_\ell (c_{\ell h} - c_{\ell\ell}) \right) \frac{(c_{h\ell} - c_{hh})}{f_\ell(c_{\ell\ell} - c_{\ell h})} + g_\ell (u_{h\ell} - c_{h\ell}) < 0 \\
\implies & \left(g_h \varphi_h - f_\ell g_\ell (c_{\ell h} - c_{\ell\ell}) \right) \frac{(c_{h\ell} - c_{hh})}{f_\ell(c_{\ell\ell} - c_{\ell h})} + g_\ell (u_{h\ell} - c_{h\ell}) < 0 \\
\implies & \left(\frac{\frac{1}{f_\ell} \varphi_h}{(c_{\ell h} - c_{\ell\ell}) \frac{g_\ell}{g_h}} - 1 \right) + \frac{(u_{h\ell} - c_{h\ell})}{c_{hh} - c_{h\ell}} < 0
\end{aligned}$$

Since the final term is positive we have that

$$\left(\frac{\frac{1}{f_\ell} \varphi_h}{(c_{\ell h} - c_{\ell\ell}) \frac{g_\ell}{g_h}} - 1 \right) < 0$$

as desired, which verifies that $\frac{\beta_{\text{IC}}}{f_h} \geq 1$.

Case 2b: Now consider the case $\frac{1}{f_\ell(c_{\ell h} - c_{\ell\ell})} \varphi_h > \frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})}$ and $k = \frac{f_h(c_{hh} - c_{h\ell})}{f_\ell(c_{\ell h} - c_{\ell\ell})} > 1$.

We will show that in this case the optimal allocation is $(0, 1, 1, x_{hh})$ where $x_{hh} \in [0, 1]$ is determined by either (S-IC-hl) binding (Case 2b (i)) or by (S-IR-h) binding (Case 2b (ii)). Observe again that only one of the constraints (S-IR-h) and (S-IC-hl) (and hence seller monotonicity) can be binding but not both. Notice, that given allocation $(0, 1, 1, x_{hh})$ seller monotonicity (S-MON) binding is equivalent to:

$$f_\ell(c_{\ell\ell} - c_{\ell h}) + f_h(c_{h\ell} - c_{hh})(x_{hh} - 1) = 0.$$

Case 2b(i): The solution to 2b(i) has seller monotonicity binding and seller IR slack. Therefore $\sigma_{\text{IR}} = 0$ by complementary slackness and as a result we know that $\frac{\beta_{\text{IC}}}{f_h} = 1$. Moreover, since our candidate solution is $(0, 1, 1, x_{hh})$, by complementary slackness we have $\eta_{\ell\ell}, \eta_{hh} = 0$ and that the dual equations that correspond to $x_{h\ell}, x_{\ell h}, x_{hh}$ must bind in the dual. Plugging this into the dual constraints, we need to show that there exists a dual feasible solution to

$$\left(u_{h\ell}\frac{f_h}{f_\ell} - u_{\ell\ell}\frac{1}{f_\ell} + c_{\ell\ell}\right) - \frac{1}{g_\ell}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \geq 0, \tag{$x_{\ell\ell}'$}$$

$$(-u_{h\ell} + c_{h\ell}) - \frac{1}{g_\ell}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} \leq 0, \tag{$x_{h\ell}'$}$$

$$\left(u_{hh}\frac{f_h}{f_\ell} - u_{\ell h}\frac{1}{f_\ell} + c_{\ell h}\right) + \frac{1}{g_h}(c_{\ell h} - c_{\ell\ell})\sigma'_{\text{IC}} \leq 0, \tag{$x_{\ell h}'$}$$

$$(-u_{hh} + c_{hh}) + \frac{1}{g_h}(c_{hh} - c_{h\ell})\sigma'_{\text{IC}} = 0, \tag{x_{hh}'}$$

$$\sigma'_{\text{IC}} \geq 0.$$

Observe that from x'_{hh} we obtain $\sigma'_{\text{IC}} = g_h(u_{hh} - c_{hh})/(c_{hh} - c_{h\ell})$. This satisfies $(x'_{\ell h})$ by our assumption. Observe that $x'_{h\ell}$ is trivially satisfied.

Finally, it is easy to check that $(x'_{\ell\ell})$ is also satisfied. Indeed,

$$-\frac{1}{f_\ell}\varphi_\ell - \frac{g_h(c_{\ell h} - c_{\ell\ell})(u_{hh} - c_{hh})}{g_\ell(c_{hh} - c_{h\ell})} \geq -\frac{1}{f_\ell}\varphi_\ell - \frac{g_h}{g_\ell f_\ell}\varphi_h \geq 0$$

where the first inequality follows since $\frac{1}{f_\ell(c_{\ell h} - c_{\ell\ell})}\varphi_h > \frac{(u_{hh} - c_{hh})}{(c_{hh} - c_{h\ell})}$ by assumption, and the second since $g_h\varphi_h + g_\ell\varphi_\ell < 0$ by assumption.

Case 2b(ii): Consider instead that we are $(0, 1, 1, x_{hh})$ such that (S-IR-h) binds and (S-IC-hl) is slack. Then by complementary slackness $\sigma'_{\text{IC}} = 0$, $\eta_{\ell\ell}, \eta_{hh} = 0$ and the dual equations that correspond to $x_{h\ell}, x_{\ell h}, x_{hh}$ must bind.

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{h\ell}\frac{f_h}{f_\ell} - u_{\ell\ell}\frac{1}{f_\ell} + c_{\ell\ell}\right) \geq 0, \tag{$x_{\ell\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}(-u_{h\ell} + c_{h\ell}) \leq 0, \tag{$x_{h\ell}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(u_{hh}\frac{f_h}{f_\ell} - u_{\ell h}\frac{1}{f_\ell} + c_{\ell h} + (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h}\right) - (c_{\ell h} - c_{\ell\ell})\frac{g_\ell}{g_h} \leq 0, \tag{$x_{\ell h}'$}$$

$$\frac{\beta_{\text{IC}}}{f_h}\left(-u_{hh} + c_{hh} + (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h}\right) - (c_{hh} - c_{h\ell})\frac{g_\ell}{g_h} = 0, \tag{x_{hh}'}$$

Note that the first two inequalities are satisfied by assumption ($u_{h\ell} \geq c_{h\ell}$ and $\varphi_\ell \leq 0$). Moreover, (x'_{hh}) yields:

$$\frac{\beta_{\text{IC}}}{f_h} = \frac{(c_{hh} - c_{h\ell})g_\ell}{g_\ell(c_{hh} - c_{h\ell}) - g_h(u_{hh} - c_{hh})},$$

which also satisfies $(x'_{\ell h})$. Indeed,

$$\begin{aligned}
&\frac{\beta_{\mathrm{IC}}}{f_h}\left(-\varphi_h\frac{1}{f_\ell}+(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}\right)-(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}\\
\leq&\frac{(c_{hh}-c_{h\ell})g_\ell}{g_\ell(c_{hh}-c_{h\ell})-g_h(u_{hh}-c_{hh})}\left(-\frac{(u_{hh}-c_{hh})(c_{\ell h}-c_{\ell\ell})}{(c_{hh}-c_{h\ell})}+(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}\right)-(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}\\
=&\frac{g_\ell}{g_h}(c_{\ell h}-c_{\ell\ell})-(c_{\ell h}-c_{\ell\ell})\frac{g_\ell}{g_h}=0.
\end{aligned}$$

We are left to check that $\frac{\beta_{\mathrm{IC}}}{f_h}\geq 1$. Notice that if the denominator is positive, then this is satisfied since $g_\ell(c_{hh}-c_{h\ell})-g_h(u_{hh}-c_{hh})<g_\ell(c_{hh}-c_{h\ell})$ by the gains from trade assumption for all types.

We use Farkas' lemma to show that the denominator of $\frac{\beta_{\mathrm{IC}}}{f_h}$, is positive, that is $g_\ell(c_{hh}-c_{h\ell})-g_h(u_{hh}-c_{hh})\geq 0$. Recall that we consider the case where (S-IR-h) binds and hence (S-IC-hl) doesn't. Then this means there is no solution to (here again we plugged the allocation into the constraints):

$$g_h\,(u_{hh}x_{hh}-t_{hh})+g_\ell(u_{h\ell}-t_{h\ell})\geq g_h(u_{hh}-t_{\ell h})+g_\ell(-t_{\ell\ell}),\tag{B-IC-hl}$$

$$g_h(u_{\ell h}-t_{\ell h})+g_\ell(-t_{\ell\ell})\geq 0,\tag{B-IR-l}$$

$$f_h(t_{h\ell}-c_{h\ell})+f_\ell(t_{\ell\ell})\geq f_h(t_{hh}-c_{h\ell}x_{hh})+f_\ell(t_{\ell h}-c_{\ell\ell}),\tag{S-IC-lh}$$

$$f_h(t_{hh}-c_{hh}x_{hh})+f_\ell(t_{\ell h}-c_{\ell h})\geq 0,\tag{S-IR-h}$$

for $(x_{\ell\ell},x_{h\ell},x_{\ell h},x_{hh})=(0,1,1,x_{hh})$ where x_{hh} solves

$$f_\ell(c_{\ell\ell}-c_{\ell h})+f_h(c_{h\ell}-c_{hh})(x_{hh}-1)=0.$$

Defining $\Delta=\frac{f_\ell(c_{\ell\ell}-c_{\ell h})}{f_h(c_{h\ell}-c_{hh})}$, implies $x_{hh}=1-\Delta$. Plugging in we can rewrite this as:

$$-g_hu_{hh}\Delta+g_\ell u_{h\ell}\geq g_h(t_{hh}-t_{\ell h})+g_\ell(t_{h\ell}-t_{\ell\ell}),\tag{A}$$

$$g_hu_{\ell h}\geq g_ht_{\ell h}+g_\ell t_{\ell\ell},\tag{B}$$

$$-f_hc_{h\ell}\Delta+f_\ell c_{\ell\ell}\geq f_h(t_{hh}-t_{h\ell})+f_\ell(t_{\ell h}-t_{\ell\ell}),\tag{C}$$

$$-f_hc_{hh}(1-\Delta)-f_\ell c_{\ell h}\geq -f_ht_{hh}-f_\ell t_{\ell h}.\tag{D}$$

By the Farkas Lemma, either there exists a solution to the system above or to the Farkas alternative below, but not both:

$$(-g_hu_{hh}\Delta+g_\ell u_{h\ell})A+g_hu_{\ell h}B-(f_hc_{h\ell}\Delta-f_\ell c_{\ell\ell})C-(f_hc_{hh}(1-\Delta)+f_\ell c_{\ell h})D<0,$$

$$g_hA+f_hC-f_hD=0,\tag{t_{hh}}$$

$$-g_hA+g_hB+f_\ell C-f_\ell D=0,\tag{$t_{\ell h}$}$$

$$g_\ell A-f_hC=0,\tag{$t_{h\ell}$}$$

$$-g_\ell A+g_\ell B-f_\ell C=0,\tag{$t_{\ell\ell}$}$$

$$A,B,C,D\geq 0.$$

Observe that $(t_{h\ell})$ implies that:

$$C=\frac{g_\ell}{f_h}A.$$

Plugging into ($t_{\ell\ell}$),

$$g_\ell B = g_\ell A + f_\ell \frac{g_\ell}{f_h} A,$$
$$\implies B = \frac{1}{f_h} A.$$

Plugging into (t_{hh}),

$$f_h D = g_h A + g_\ell A$$
$$\implies D = \frac{1}{f_h} A.$$

It is easy to verify that this satisfies ($t_{\ell h}$).

Finally note that

$$
\begin{aligned}
&(-g_h u_{hh}\Delta + g_\ell u_{h\ell})A + g_h u_{\ell h} B - (f_h c_{h\ell}\Delta - f_\ell c_{\ell\ell})C - (f_h c_{hh}(1-\Delta) + f_\ell c_{\ell h})D \\
=&(-g_h u_{hh}\Delta + g_\ell u_{h\ell})A + g_h u_{\ell h}\frac{A}{f_h} - (f_h c_{h\ell}\Delta - f_\ell c_{\ell\ell})\frac{g_\ell}{f_h}A - (f_h c_{hh}(1-\Delta) + f_\ell c_{\ell h})\frac{A}{f_h} \\
=&\left(-g_h u_{hh}\Delta + g_\ell u_{h\ell} + \frac{g_h}{f_h}u_{\ell h} - g_\ell c_{h\ell}\Delta + \frac{f_\ell g_\ell}{f_h}c_{\ell\ell} - c_{hh}(1-\Delta) - \frac{f_\ell}{f_h}c_{\ell h}\right) A \\
=&\left(\frac{g_h}{f_h}\varphi_h + g_h(u_{hh} - c_{hh})(1-\Delta) - \frac{g_\ell f_\ell}{f_h}(c_{\ell h} - c_{\ell\ell}) + g_\ell(u_{h\ell} - c_{h\ell}) + g_\ell c_{h\ell}(1-\Delta) - g_\ell c_{hh}(1-\Delta)\right) A \\
=&\left(\frac{g_h}{f_h}\varphi_h + g_h(u_{hh} - c_{hh})(1-\Delta) + g_\ell(u_{h\ell} - c_{h\ell}) + \frac{f_\ell g_\ell}{f_h}(c_{\ell\ell} - c_{\ell h}) + g_\ell(c_{h\ell} - c_{hh})(1-\Delta)\right) A \\
=&\left(\frac{g_h}{f_h}\varphi_h + g_h(u_{hh} - c_{hh})(1-\Delta) + g_\ell(u_{h\ell} - c_{h\ell}) + g_\ell(c_{h\ell} - c_{hh})\Delta + g_\ell(c_{h\ell} - c_{hh})(1-\Delta)\right) A \\
=&\left(\frac{g_h}{f_h}\varphi_h + g_h(u_{hh} - c_{hh})(1-\Delta) + g_\ell(u_{h\ell} - c_{h\ell}) + g_\ell(c_{h\ell} - c_{hh})\right) A \\
=&\left(\left(\frac{g_h}{f_h}\varphi_h - g_h(u_{hh} - c_{hh})\Delta\right) + g_\ell(u_{h\ell} - c_{h\ell}) + (g_h(u_{hh} - c_{hh}) - g_\ell(c_{h\ell} - c_{hh}))\right) A < 0
\end{aligned}
$$

Where the third equation from below follows by using our definition of Δ. The inequality follows from analyzing the first factor: the first term here is positive by assumption in this case (after plugging in Δ), the second term is positive by the assumption of gains from trade for all types. Therefore for the overall sum to be negative, it must be that the final term $g_h(u_{hh} - c_{hh}) - g_\ell(c_{h\ell} - c_{hh})$ is negative, as desired.

This completes the proof of the theorem.

■

A.3. *Equilibrium*

We now show that the optimal allocation characterized in Theorem 4 and Theorem 5can be supported as equilibrium. Here strategies are described as:

(1) Seller strategy: Both types of seller announce the same ex-ante optimal mechanism.
(2) Buyer strategy:
   - If the seller announces the ex-ante optimal mechanism, then the buyer's beliefs about the seller's type equals their prior belief and the buyer reports their type truthfully in the mechanism.
   - If instead the seller announces *any* other mechanism, the buyer believes that the seller is of the low type, and best responds given that belief.

THEOREM 6. *Consider the ex-ante optimal mechanism for the seller as described in Theorems 4 and 5. This mechanism can be supported as an equilibrium of the informed principal game.*

PROOF. Note that upon a deviation from the optimal mechanism, the buyer's belief is that the seller is of the low type.

We use the notation $\pi^{EO}(r;\omega)$, to denote the interim profit of the seller in the ex-ante optimal mechanism if her type is ω, she reports r, and the buyer plays the equilibrium strategy. We denote by $\pi^{CK}(\omega)$ the optimal expected profit of the seller of type ω if her type were common knowledge. Finally, recall that $\pi_\omega := \max\{f_h[u_{\theta_h\omega_\ell} - c_\omega], u_{\theta_\ell\omega_\ell} - c_\omega, 0\}$ is the highest possible profit that a seller of type ω can obtain if the buyer's belief assigns probability one to the seller's type being ω_ℓ. Observe that this eis the maximal possible profit for a seller of type ω if she deviates.

As discussed in the paper, we can set up a linear program that characterizes the seller-optimal equilibrium by replacing the seller IR constraint in (OPT) with the constraint that the payoff from the mechanism for each seller type ω must be at least π_ω—the maximal profit that this type could obtain from deviating. Formally, the relevant IR constraints for each type ω of the seller are replaced by the following no deviation constraints:

$$\sum_b f_b(t_{b\omega} - c_{b\omega}\cdot x_{b\omega}) \geq \pi_\omega \qquad \forall\omega\in\Omega. \tag{N-DEV}$$

Our approach will be to take as fixed the optimal allocation x that solves the ex-ante optimal mechanism design problem of the seller (OPT), and show that there exist transfers that satisfy the seller's no-deviation constraints; i.e., not just an outside option of 0 but the (possibly) higher value of π_ω.

We will verify equilibrium by considering three separate cases:

Case 1: $\varphi_\ell < 0$, $\varphi_h > 0$. In this case note that $\pi_\ell = f_h(u_{h\ell} - c_{h\ell})$. Further, $\pi_h = \max\{f_h(u_{h\ell} - c_{hh}), 0\}$, since $u_{\ell\ell} - f_h u_{h\ell} - f_\ell c_{\ell h} \leq u_{\ell\ell} - f_h u_{h\ell} - f_\ell u_{\ell\ell} - \varphi_\ell < 0$.

Notice that, for each of the cases in Theorem 4, the optimal allocation rule prescribes allocation $(x_{\ell\ell}, x_{h\ell}) = (0,1)$ for the low-type seller. Further the transfer in the ex-ante optimal mechanism for the low-type seller must satisfy $f_h t_{h\ell} + f_\ell t_{\ell\ell} \geq f_h u_{h\ell}$ and hence $\pi^{EO}(\omega_\ell;\omega_\ell) = (f_h t_{h\ell} + f_\ell t_{\ell\ell}) - f_h c_{h\ell} \geq f_h(u_{h\ell} - c_{h\ell}) = \pi_\ell$. That is, the low-type seller does not want to deviate.. Moreover, the high type seller does not want to deviate (since (S-IC-hl) is satisfied) and this is at least as attractive a deviation. Indeed, given the optimal allocation prescribes $(x_{\ell\ell}, x_{h\ell}) = (0, 1)$ for the low-type seller, by (S-IC-hl) we obtain

$$f_h(t_{hh} - c_{hh}x_{hh}) + f_\ell(t_{\ell h} - c_{\ell h}x_{\ell h}) \geq f_h(t_{h\ell}) + f_\ell(t_{\ell\ell} - c_{\ell h}) \geq f_h u_{h\ell} - f_h c_{hh} = \pi_h.$$

Case 2: $\varphi_\ell > 0$, $\varphi_h < 0$. This is the case considered in Theorem 4. In this case note that:

$$\pi_\ell = u_{\ell\ell} - f_h c_{h\ell} - f_\ell c_{\ell\ell},$$
$$\pi_h = \max\{0, f_h(u_{h\ell} - c_{hh}), u_{\ell\ell} - f_h c_{hh} - f_\ell c_{\ell h}\}.$$

Case 2a: $\pi_h = f_h(u_{h\ell} - c_{hh})$.

First observe that in this subcase it must be that $u_{h\ell} - c_{hh} \geq 0$. Observe that this implies

$$\begin{aligned}
& g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} \\
\geq & g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - u_{h\ell} && (\text{ since } u_{h\ell} - c_{hh} \geq 0) \\
= & g_h(u_{hh} - u_{h\ell}) + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - g_\ell u_{h\ell} \\
= & g_h(u_{hh} - u_{h\ell}) + \frac{g_\ell}{f_h} \varphi_\ell \\
\geq & 0.
\end{aligned}$$

and hence by Claim 1 the optimal allocation rule to the ex-ante optimal mechanism is $(x_{\ell\ell}, x_{\ell h}, x_{h\ell}, x_{hh}) = (1, 0, 1, 1)$.

Plugging $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$ into (B-IC-hl), (B-IR-l), (S-IC-lh) and replacing the seller's IR constraints with the respective no deviation constraints (N-DEV) yields the following system of inequalities (after collecting terms). The tags of the inequalities represent the corresponding variable in the Farkas alternative.

$$g_h u_{hh} \geq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A}$$
$$g_\ell u_{\ell\ell} \geq g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B}$$
$$-f_\ell c_{\ell\ell} \geq -f_h(t_{h\ell} - t_{hh}) - f_\ell(t_{\ell\ell} - t_{\ell h}), \tag{C}$$
$$-f_h u_{h\ell} \geq -f_h t_{hh} - f_\ell t_{\ell h}, \tag{D}$$
$$-u_{\ell\ell} \geq -f_h t_{h\ell} - f_\ell t_{\ell\ell}. \tag{E}$$

By the Farkas' Lemma, either there exists a solution to the system above or to the Farkas alternative below, but not both:

$$g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h u_{h\ell} D - u_{\ell\ell} E < 0,$$
$$g_h A + f_h C - f_h D = 0, \tag{t_{hh}}$$
$$-g_h A + g_h B + f_\ell C - f_\ell D = 0, \tag{$t_{\ell h}$}$$
$$g_\ell A - f_h C - f_h E = 0, \tag{$t_{h\ell}$}$$
$$-g_\ell A + g_\ell B - f_\ell C - f_\ell E = 0, \tag{$t_{\ell\ell}$}$$
$$A, B, C, D, E \geq 0.$$

We will show that for any non-negative solution to (t_{hh}, $t_{\ell h}$, $t_{h\ell}$, $t_{\ell\ell}$), we must have that $g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h u_{h\ell} D - u_{\ell\ell} E \geq 0$.

Now, note that f_ℓ times equation (t_{hh}) minus f_h times ($t_{\ell h}$) implies

$$f_h B = A.$$

Similarly, adding equation (t_{hh}) and equation ($t_{\ell h}$), we have

$$C - D + g_h B = 0. \tag{13}$$

Adding equation ($t_{h\ell}$) and equation ($t_{\ell\ell}$), we have

$$C + E = g_\ell B. \tag{14}$$

Subtracting equation (13) from (14) we have

$$D + E = B.$$

Finally, multiplying equation (13) by g_ℓ, and equation (14) by g_h and adding we have

$$C = g_\ell D - g_h E.$$

Now we can evaluate

$$g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h u_{h\ell} D - u_{\ell\ell} E.$$

to that end, note that from above we have that $A = f_h(D+E)$, $B = D + E$, $C = g_\ell D - g_h E$. Substituting all these in, we obtain

$$\begin{aligned}
& g_h u_{hh} A + g_\ell u_{\ell\ell} B - f_\ell c_{\ell\ell} C - f_h u_{h\ell} D - u_{\ell\ell} E \\
=& g_h u_{hh} f_h (D+E) + g_\ell u_{\ell\ell}(D+E) - f_\ell c_{\ell\ell}(g_\ell D - g_h E) - f_h u_{h\ell} D - u_{\ell\ell} E, \\
=& D\left(g_h u_{hh} f_h + g_\ell u_{\ell\ell} - f_\ell c_{\ell\ell} g_\ell - f_h u_{h\ell}\right) + E\left(g_h u_{hh} f_h - g_h u_{\ell\ell} + f_\ell c_{\ell\ell} g_h\right), \\
=& D\left(g_h f_h (u_{hh} - u_{h\ell}) + g_\ell \varphi_\ell\right) + E\left(g_h u_{hh} f_h - g_h u_{\ell\ell} + f_\ell c_{\ell\ell} g_h\right) \\
=& D\left(g_h f_h (u_{hh} - u_{h\ell}) + g_\ell \varphi_\ell\right) + E\left(g_h f_h (u_{hh} - u_{h\ell}) - g_h \varphi_\ell\right) \\
=& g_h f_h (u_{hh} - u_{h\ell})(D+E) + \varphi_\ell (g_\ell D - g_h E) \\
=& g_h f_h (u_{hh} - u_{h\ell})(D+E) + \varphi_\ell C. \geq 0,
\end{aligned}$$

since $u_{hh} - u_{h\ell} \geq 0$ (monotonicity) and $\varphi_\ell \geq 0$ by assumption, and moreover $C, D, E \geq 0$ must be non-negative in a feasible solution. This shows that the Farkas' alternative has no solution and hence there exists feasible transfers that satisfy the above system of constraints (A) – (E).
Case 2b: $\pi_h = u_{\ell\ell} - f_h c_{hh} - f_\ell c_{\ell h}$.

Observe that this implies $u_{\ell\ell} - f_h c_{hh} - f_\ell c_{\ell h} \geq 0$ from which it follows that

$$\begin{aligned}
g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} =& g_h (u_{hh} - c_{hh}) + \frac{g_\ell}{f_h}\left(u_{\ell\ell} - f_\ell c_{\ell\ell} - f_h c_{hh}\right) \\
\geq& g_h (u_{hh} - c_{hh}) + \frac{g_\ell}{f_h}\left(u_{\ell\ell} - f_\ell c_{\ell h} - f_h c_{hh}\right) \qquad (\text{ since } c_{\ell h} \geq c_{\ell\ell}) \\
\geq& 0.
\end{aligned}$$

Here the last inequality follows since $u_{hh} - c_{hh} \geq 0$ by assumption and $(u_{\ell\ell} - f_\ell c_{\ell h} - f_h c_{hh}) \geq 0$ in this subcase. Now, from Claim 1 we know that the optimal allocation in the ex-ante optimal mechanism is $(x_{\ell\ell}, x_{\ell h}, x_{h\ell}, x_{hh}) = (1, 0, 1, 1)$.

Plugging the allocation $x_{hh} = x_{h\ell} = x_{\ell\ell} = 1, x_{\ell h} = 0$ into (B-IC-hl), (B-IR-l), (S-IC-lh) and replacing the seller's IR constraints with the respective no-deviation constraints (N-DEV) yields the following system of inequalities (after collecting terms):

$$g_h u_{hh} \geq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A}$$

$$g_\ell u_{\ell\ell} \geq g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B}$$

$$f_h(t_{h\ell} - t_{hh}) + f_\ell(t_{\ell\ell} - t_{\ell h}) \geq f_\ell c_{\ell\ell}, \tag{C}$$

$$f_h t_{hh} + f_\ell t_{\ell h} \geq u_{\ell\ell} - f_\ell c_{\ell h}, \tag{D}$$

$$f_h t_{h\ell} + f_\ell t_{\ell\ell} \geq u_{\ell\ell}. \tag{E}$$

It is therefore sufficient if we show the following system has a solution:

$$g_h u_{hh} = g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A'}$$

$$g_\ell u_{\ell\ell} = g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B'}$$

$$f_h(t_{h\ell} - t_{hh}) + f_\ell(t_{\ell\ell} - t_{\ell h}) \geq f_\ell c_{\ell\ell}, \tag{C}$$

$$f_h t_{hh} + f_\ell t_{\ell h} \geq u_{\ell\ell} - f_\ell c_{\ell h}, \tag{D}$$

$$f_h t_{h\ell} + f_\ell t_{\ell\ell} \geq u_{\ell\ell}. \tag{E}$$

Adding (A') and (B') we have:

$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} + g_\ell u_{\ell\ell}. \tag{F}$$

Adding f_ℓ times (B') and f_h times (F) yields:

$$f_\ell(g_h t_{\ell h} + g_\ell t_{\ell\ell}) + f_h(g_h t_{hh} + g_\ell t_{h\ell}) = g_\ell u_{\ell\ell} + f_h g_h u_{hh}. \tag{G}$$

Note that from (D) we have that $f_h t_{hh} + f_\ell t_{\ell h} \geq u_{\ell\ell} - f_\ell c_{\ell h}$. Multiplying (D) by g_h, subtracting from (G), and dividing by g_ℓ throughout, we obtain:

$$f_\ell t_{\ell\ell} + f_h t_{h\ell} \leq u_{\ell\ell} + \frac{g_h}{g_\ell}(f_h u_{hh} - u_{\ell\ell} + f_\ell c_{\ell h}) \tag{H}$$

Since by assumption $\varphi_h < 0$, it follows that:

$$\begin{aligned} & u_{\ell\ell} - f_h u_{hh} - f_\ell c_{\ell h} < u_{\ell h} - f_h u_{hh} - f_\ell c_{\ell h} < 0 \\ \implies & -u_{\ell\ell} + f_h u_{hh} + f_\ell c_{\ell h} > 0. \end{aligned} \tag{15}$$

In short, consider the following system:

$$g_h t_{\ell h} + g_\ell t_{\ell\ell} = g_\ell u_{\ell\ell} \tag{B'}$$

$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} + g_\ell u_{\ell\ell} \tag{F}$$

$$f_h t_{hh} + f_\ell t_{\ell h} = u_{\ell\ell} - f_\ell c_{\ell h} + \delta \tag{D'}$$

$$f_\ell t_{\ell\ell} + f_h t_{h\ell} = u_{\ell\ell} + \frac{g_h}{g_\ell}(f_h u_{hh} - u_{\ell\ell} + f_\ell c_{\ell h}) - \frac{g_h}{g_\ell}\delta \tag{H'}$$

where δ solves

$$\frac{g_h}{g_\ell}(f_h u_{hh} - u_{\ell\ell} + f_\ell c_{\ell h}) - \frac{g_h}{g_\ell}\delta + f_\ell c_{\ell h} - \delta = f_\ell c_{\ell\ell}$$
$$\implies \frac{1}{g_\ell}\delta = \frac{g_h}{g_\ell}(f_h u_{hh} - u_{\ell\ell} + f_\ell c_{\ell h}) + f_\ell(c_{\ell h} - c_{\ell\ell}).$$

Note that δ therefore is positive by construction (cf (15) and $c_{\ell h} \geq c_{\ell\ell}$). It is straightforward to verify that a solution to this system will satisfy (A–E). Furthermore, notice that a solution to the first three equations (B′, F, D′) satisfies the fourth. To see this, notice that $f_\ell(B') + f_h(F) - g_h(D') = g_\ell(H')$. Finally, notice that the system of the first three equations must have a solution (3 equations in 4 variables).

Case 2c: $\pi_h = 0$.

Observe that in this case, the outside option of the high type in the mechanism design problem is 0, i.e., the same as the IR constraint of the high-type in the ex-ante mechanism design problem.

First suppose the allocation rule in the ex-ante optimal is $(x_{\ell\ell}, x_{\ell h}, x_{h\ell}, x_{hh}) = (1, 0, 1, 1)$.

In this case we need to show there exists a solution to:

$$g_h u_{hh} \geq g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A}$$
$$g_\ell u_{\ell\ell} \geq g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B}$$
$$f_h(t_{h\ell} - t_{hh}) + f_\ell(t_{\ell\ell} - t_{\ell h}) \geq f_\ell c_{\ell\ell}, \tag{C}$$
$$f_h t_{hh} + f_\ell t_{\ell h} \geq f_h c_{hh}, \tag{D}$$
$$f_h t_{h\ell} + f_\ell t_{\ell\ell} \geq u_{\ell\ell}. \tag{E}$$

It is therefore sufficient if we show the following system has a solution:

$$g_h u_{hh} = g_h(t_{hh} - t_{\ell h}) + g_\ell(t_{h\ell} - t_{\ell\ell}), \tag{A′}$$
$$g_\ell u_{\ell\ell} = g_h t_{\ell h} + g_\ell t_{\ell\ell}, \tag{B′}$$
$$f_h(t_{h\ell} - t_{hh}) + f_\ell(t_{\ell\ell} - t_{\ell h}) \geq f_\ell c_{\ell\ell}, \tag{C}$$
$$f_h t_{hh} + f_\ell t_{\ell h} \geq f_h c_{hh}, \tag{D}$$
$$f_h t_{h\ell} + f_\ell t_{\ell\ell} \geq u_{\ell\ell}. \tag{E}$$

Adding (A') and (B') we have :

$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} + g_\ell u_{\ell\ell}. \tag{F}$$

Adding f_ℓ times (B) and f_h times (F) yields:

$$f_\ell(g_h t_{\ell h} + g_\ell t_{\ell\ell}) + f_h(g_h t_{hh} + g_\ell t_{h\ell}) = g_\ell u_{\ell\ell} + f_h g_h u_{hh}. \tag{G}$$

Note that from (D) we have that $f_h t_{hh} + f_\ell t_{\ell h} \geq f_h c_{hh}$. Multiplying (D) by g_h subtracting from (G), and dividing by g_ℓ throughout, we see that:

$$f_\ell t_{\ell\ell} + f_h t_{h\ell} \leq u_{\ell\ell} + \frac{g_h}{g_\ell} f_h(u_{hh} - c_{hh}) \tag{H}$$

In short, consider the following system:

$$g_h t_{\ell h} + g_\ell t_{\ell\ell} = g_\ell u_{\ell\ell} \tag{B'}$$

$$g_h t_{hh} + g_\ell t_{h\ell} = g_h u_{hh} + g_\ell u_{\ell\ell} \tag{F}$$

$$f_h t_{hh} + f_\ell t_{\ell h} = f_h c_{hh} + \delta \tag{D'}$$

$$f_\ell t_{\ell\ell} + f_h t_{h\ell} = u_{\ell\ell} + \frac{g_h}{g_\ell} f_h (u_{hh} - c_{hh}) - \frac{g_h}{g_\ell}\delta \tag{H'}$$

where δ solves

$$\begin{aligned} & u_{\ell\ell} + \frac{g_h}{g_\ell} f_h (u_{hh} - c_{hh}) - \frac{g_h}{g_\ell}\delta - f_h c_{hh} - \delta = f_\ell c_{\ell\ell} \\ \Longrightarrow\ & \frac{1}{g_\ell}\delta = u_{\ell\ell} + \frac{g_h}{g_\ell} f_h u_{hh} - \frac{1}{g_\ell} f_h c_{hh} - f_\ell c_{\ell\ell}. \\ \Longrightarrow\ & \delta = g_\ell u_{\ell\ell} + g_h f_h u_{hh} - f_h c_{hh} - f_\ell g_\ell c_{\ell\ell}. \end{aligned}$$

Recall that from Claim 1 we know that the fact the ex-ante optimal mechanism involves allocation $(x_{\ell\ell}, x_{\ell h}, x_{h\ell}, x_{hh}) = (1, 0, 1, 1)$ if and only if:

$$\begin{aligned} & g_h u_{hh} + \frac{g_\ell}{f_h} u_{\ell\ell} - \frac{f_\ell g_\ell}{f_h} c_{\ell\ell} - c_{hh} \geq 0. \\ \Longrightarrow\ & g_h f_h u_{hh} + g_\ell u_{\ell\ell} - f_\ell g_\ell c_{\ell\ell} - f_h c_{hh} \geq 0 \\ \Longrightarrow\ & \delta \geq 0 \end{aligned}$$

By construction, a solution to this system of equations (B', F, D', H') will satisfy (A–E).

Further, note that a solution to the first three equations (B', F, D') satisfies the fourth (H'). Finally note that the system of the first three equations must have a solution (3 equations in 4 variables).

Finally suppose $(x_{\ell\ell}, x_{\ell h}, x_{h\ell}, x_{hh}) = (1, 0, 1, x)$ for some $x < 1$. Note that $x < 1$ implies that the allocation rule $(1, 0, 1, 1)$ does not have a transfer rule that satisfies (S-IR-h).

Since the high-type of the seller also makes no profit (since (S-IR-h) must bind in the optimal mechanism here), it follows that all the gains from trade for the seller must accrue to the low-type. Since the low-type is already offering the optimal allocation rule if they were to deviate, they must be making higher profits than that.

■

REFERENCES

AKERLOF, G. A. (1970): "The Market for Lemons: Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84(3), 488–500.

BERGEMANN, D., B. BROOKS, AND S. MORRIS (2015): "The Limits of Price Discrimination," *American Economic Review*, 105(3), 921–57.

BERGEMANN, D., AND S. MORRIS (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), 44–95.

BRUNNERMEIER, M. K., R. LAMBA, AND C. SEGURA-RODRIGUEZ (2021): "Inverse Selection," *Available at SSRN 3584331*.

CALZOLARI, G., AND A. PAVAN (2006): "On the Optimality of Privacy in Sequential Contracting," *Journal of Economic theory*, 130(1), 168–204.

CONITZER, V., C. R. TAYLOR, AND L. WAGMAN (2012): "Hide and Seek: Costly Consumer Privacy in a Market with Repeat Purchases," *Marketing Science*, 31(2), 277–292.

CUMMINGS, R., K. LIGETT, M. M. PAI, AND A. ROTH (2016): "The Strange Case of Privacy in Equilibrium Models," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, pp. 659–659.

DWORK, C., AND A. ROTH (2014): "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, 9(3–4), 211–407.

KAMENICA, E. (2019): "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 11, 249–272.

KARTIK, N., AND W. ZHONG (2023): "Lemonade from Lemons: Information Design and Adverse Selection," Discussion paper, Working Paper.

KOESSLER, F., AND V. SKRETA (2016): "Informed Seller with Taste Heterogeneity," *Journal of Economic Theory*, 165, 456–471.

MASKIN, E., AND J. TIROLE (1990): "The Principal-Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, pp. 379–409.

——— (1992): "The Principal-Agent Relationship with an Informed Principal, II: Common Values," *Econometrica*, pp. 1–42.

MYERSON, R. B. (1983): "Mechanism Design by an Informed Principal," *Econometrica*, 51(6), 1767–1797.

MYLOVANOV, T., AND T. TRÖGER (2014): "Mechanism Design by an Informed Principal: Private Values with Transferable Utility," *The Review of Economic Studies*, 81(4), 1668–1707.

NISHIMURA, T. (2022): "Informed Principal Problems in Bilateral Trading," *Journal of Economic Theory*, 204, 105498.

ROESLER, A.-K., AND B. SZENTES (2017): "Buyer-Optimal Learning and Monopoly Pricing," *American Economic Review*, 107(7), 2072–80.

TAYLOR, C. R. (2004): "Consumer Privacy and the Market for Customer Information," *RAND Journal of Economics*, pp. 631–650.