Robust Trade in Lemons Markets

Gabriel Carroll, Stanford University

gdc@stanford.edu

December 31, 2013

Abstract

A buyer and seller have the opportunity to exchange an indivisible good at a prespecified price. Each agent may be imperfectly informed, in an arbitrary way, about both his own value for the good and the other agent's value. An observer knows the joint distribution of the two agents' values, but does not know their information structure. We determine what lower bounds the observer can confidently predict for the expected gains from trade that can be realized in equilibrium. In particular, we show that the worst-case information structure — that minimizes the realizable gains from trade — involves no information asymmetries.

Thanks to (in random order) Paul Milgrom, Alp Simsek, Stephan Lauermann, Alex Wolitzky, Matthew Jackson, Anton Tsoy, Nathan Hendren, Daron Acemoglu, Jean Tirole, and Richard Holden for helpful comments.

1 Introduction

Suppose that a buyer and seller meet to exchange a single, indivisible good at a known price. In a classical economic model — say, an exchange economy — the parties' values for the good being traded are commonly known. If, for example, the good is worth 6 to the buyer and 4 to the seller, and they are able to trade at a price of 5, then an observer can confidently predict that the parties will be able to realize the gains from trade of 2.

All well and good. Now suppose that the values are not commonly known. Suppose instead that they are distributed according to the following prior:

- with probability 80% the world is in a "normal" state, in which the object has value 6 to the buyer and 4 to the seller, as before;
- there is also 10% chance of a "good" state, in which the parties' values for the object are 8 and 6 respectively;
- there is a 10% chance of a "bad" state, in which the parties' values are 4 and 2.

Suppose, moreover, that the observer knows this distribution, but does not know what information the trading parties have about the state. They may both be fully informed; or there may be a lemons market as in Akerlof [1], where the seller knows the state while the buyer is uninformed; or the buyer may know the state and the seller may be uninformed; or both parties may receive independent signals of the state, with each signal having a 1/2 probability of being correct and a 1/2 probability of being uninformative noise; or perhaps the information structure is something much more complicated. In general, it will not be possible to realize all the gains from trade. Indeed, a long tradition in information economics has pointed out how information asymmetries can lead to breakdown of trade, starting with Akerlof [1] and continuing to more recent contributions that stress the importance of higher-order beliefs [17, 2, 14].

But can the observer predict at least *some* probability of trade? Indeed she can. In this example, even without knowing the information structure, the observer can predict that the parties will be able to trade with probability at least 60%. More precisely, no matter what the information structure is, as long as the buyer and seller share a common prior over it, the resulting Bayesian game between them has an equilibrium in which at least 60% of the gains from trade are realized (in expectation); and this 60% prediction is sharp.

This paper will show, more generally, how the observer who knows only the distribution of traders' values can compute the sharpest possible prediction for attainable gains from trade. We will also describe the information structure that makes this prediction sharp. Perhaps surprisingly, this worst-case information structure does not involve asymmetric information. Instead, both parties receive the same signal: either a "high-value" signal (in which case the seller does not want to trade at the posted price, because his expected value from keeping the good is higher); or a "low-value signal" (such that the buyer does not want to trade); or a "normal" signal. (The result that the worst case involves symmetric information is specific to the posted-price mechanism that we assume; this will be further discussed in the conclusion.)

On some level, our characterization of the worst case is unsurprising: if trade fails, it should be either because the seller expects his value is too high or the buyer expects his value is too low. But the result is not trivial because of contagion effects through the interaction of the two parties' information: as in any situation of adverse selection, in equilibrium each trader should decide to trade not only based on his interim belief about his value but based on the information content of the other party's willingness to trade. Consequently, for any particular information structure, describing equilibrium behavior can be nontrivial.

Here is a simple example illustrating this point; one can readily cook up more complicated examples. We stick with the 80–10–10 distribution of values described above. Suppose that the information structure is as follows: The buyer's signal η_B may take one of three possible realizations, which we call A, B, C; the seller's signal η_S may take on realizations D, E, F. The following table shows the probability of each pair of signals, as well as the buyer's and seller's values for the goods corresponding to each possible pair of signals. (In this example, each possible pair of signals can occur for only one state of the world, but our general model will not assume this.)

$\eta_B \backslash \eta_S$	D	E	F
A	0.36 $6, 4$	0.03 $4, 2$	0.40 $6, 4$
B	0.04 $6, 4$	0.05 $4, 2$	
C		0.10 $8, 6$	0.02 $4, 2$

Table 1: Joint distribution of signals and values

The buyer and seller observe their respective signals, and then decide whether to agree to trade; if both agree, they trade at the price of 5.

In this example, if the seller receives signal D or F, then for sure he benefits from trading, so we may as well assume he agrees to trade. Then, if the buyer receives signal A, his expected gain from agreeing to trade is positive (although its exact value depends how the seller with signal E behaves), so the buyer with signal A agrees as well.

Does the seller with signal E agree to trade? If he does, then we can check that the buyer with signal B prefers not to trade, while the buyer with signal C prefers to trade. Given that the buyer trades under signals A and C but not B, then the seller with signal

E earns negative gains from trade, so prefers not to trade.

On the other hand, if the seller does not trade under signal E, then the buyer prefers to trade under signal B and not C. In this case, the seller's best reply under signal E is to trade.

So it must be in equilibrium that the seller mixes under signal E. With a little more calculation, we can check that the equilibrium is as follows: the buyer agrees to trade with probability $1/15$ following signal B, and probability 1 for signal C; and the seller agrees with probability $4/5$ under signal B. The resulting probability of both parties agreeing to trade is $341/375 \approx 0.91$.

Examples of this sort suggest that it might be possible to devise more complicated information structures, perhaps using email-game-like constructions [17] to create ripple effects across information sets, so that ultimately one party or the other is unwilling to trade almost all the time. Indeed, one can easily give such constructions to make each party *sometimes* unwilling to trade even in states where his ex-post gains are very high. But our results show that such contagion cannot prevent trade all the time, and sharply delineate just how bad it can be, from an ex-ante view.

Now that we have sketched out our results, it is tempting to discuss interpretations and possible applications. However, it will be easiest to give this discussion clearly after having given the full statement of the model and results, and indicating their limitations. So we leave the economic interpretations to the concluding Section 5 — with apologies to any hurried readers — and instead devote the rest of this introduction to the paper's methodological contribution and its context.

The broader purpose of the analysis here is to advance exploration of what can be predicted about agents' behavior without knowing their information structure. In that respect, the exercise we perform, and the motivation, are similar to the work of Bergemann and Morris, who take a similar approach at an abstract level to general static games [5] and apply it to games with a quadratic-normal structure [6], and to Bergemann, Brooks and Morris, who perform a similar analysis in a monopoly pricing problem [4] and a first-price auction [3]. Like the latter two papers in particular, we choose a relatively simple and common form of economic interaction and explore the possible information-free predictions.

One difference between our work and the others just mentioned is that the latter explore *all* equilibria for a given information structure, whereas we focus on the best equilibrium. Indeed, in our setup, it is always an equilibrium for both parties to never agree to trade. Moreover, this equilibrium cannot always be eliminated with a simple

refinement (see Subsection 4.2). Hence if we allowed all equilibria, the observer could make no predictions about the realized gains from trade.

Because of this difference, it can be difficult to interpret our results as a positive prediction for what trades will happen, unless one accepts some optimistic equilibrium selection that is not modeled here. Otherwise, our contribution can be better thought of as exploring how much difference information can make in determining what is possible — showing limits to how much breakdown of trade can be blamed on information structure.

Another technical connection from our results is with the literature on ex-ante robustness of equilibria of complete-information games, as in Kajii and Morris [9]. Indeed, one way to look at our results is to focus on how they imply a continuity in the best equilibrium outcomes: In the classical economy outlined above, where there is probability 1 that both parties benefit from trade, all the gains from trade can be realized; our results imply that if additional states where one party or the other does not gain are introduced with small probability, it remains possible to realize most of the gains from trade. This continuity actually already follows from the Kajii-Morris results; in our complete-information game, the always-trade equilibrium is robust in their sense. But beyond the continuity statement, we give a sharp quantitative result. Our quantitative conclusion can be seen as complementary to that of Kajii and Morris in our setting: They show how to find a sharp quantitative bound on how much the complete-information equilibrium can move when new states can be introduced with arbitrary payoff structures; we restrict to a specific game and specific payoff structures, and give a corresponding bound by different techniques.

2 Model

Let's now flesh out the formal model. The buyer's and seller's values for the good, b and s, are random variables whose joint distribution is given by a probability measure μ on $\mathbb{R}^2$, with compact support. This μ describes the prior belief, shared by the buyer, seller, and the observer. We assume $b \geq s$ with probability 1: it is common knowledge that there are (weak) gains from trade. (We will discuss later the consequences of relaxing this assumption.)

We assume a very simple institution for trading. There is a known market price p, which is constant. Each of the two agents can either agree to trade at that price or decline to trade. If both agents agree, they trade, giving payoffs $b - p$ and $p - s$ to the buyer and seller respectively. If either declines, then both receive payoff 0.

We will assume that neither the buyer nor the seller is certain ex ante that trade is beneficial for him: the events $b - p < 0$ and $p - s < 0$ both have positive probability. (If either of these events has probability zero, the problem is much simpler; we will briefly address this situation later.)

Both the buyer and seller may receive information prior to trading, via an *information structure* which is unknown to the observer. We restrict to finite information structures, to avoid complications with equilibrium existence. Thus, an information structure consists of two finite sets of signals, $\mathcal{I}_B$ and $\mathcal{I}_S$, and a joint probability measure ν on $\mathbb{R}^2 \times \mathcal{I}_B \times \mathcal{I}_S$, such that the marginal of ν on the $\mathbb{R}^2$ component coincides with μ. The signals received by the two agents will be denoted by $\eta_B \in \mathcal{I}_B$ and $\eta_S \in \mathcal{I}_S$.

Any information structure induces a Bayesian game, in which the two agents observe their signals and then decide whether to agree to trade. The buyer's possible (mixed) strategies are functions $\sigma_B : \mathcal{I}_B \to [0, 1]$, denoting the probability of agreeing after each signal, and the seller's strategies are functions $\sigma_S : \mathcal{I}_S \to [0, 1]$. The expected payoffs from a strategy profile are

$$u_B(\sigma_B, \sigma_S) = \int \sigma_B(\eta_B)\sigma_S(\eta_S)(b - p)\, d\nu, \qquad u_S(\sigma_B, \sigma_S) = \int \sigma_B(\eta_B)\sigma_S(\eta_S)(p - s)\, d\nu. \tag{2.1}$$

A strategy profile (σ_B, σ_S) is a (Bayesian Nash) *equilibrium* if

$$u_B(\sigma_B, \sigma_S) \geq u_B(\sigma'_B, \sigma_S) \qquad \text{and} \qquad u_S(\sigma_B, \sigma_S) \geq u_S(\sigma_B, \sigma'_S)$$

for any alternative strategies σ'_B, σ'_S.

The observer would like to make robust predictions about the best possible equilibrium, as measured by some criterion, e.g. the highest expected surplus, or highest probability of trade. (Expected surplus is perhaps the most natural criterion, but we may as well allow for others, since it will require little extra work.) We assume the observer's objective is represented by some bounded, measurable function of b, s, call it $w(b, s)$: the observer gets $w(b, s)$ when trade occurs and 0 otherwise. Thus, the observer's criterion is

$$W(\sigma_B, \sigma_S) = \int \sigma_B(\eta_B)\sigma_S(\eta_S)w(b, s)\, d\nu.$$

For example, if we define $w(b, s) = b - s$ then this captures the expected gains from trade realized in equilibrium; if $w(b, s) = 1$ then we have the probability of trade. Other criteria might express the observer's placing more importance on trade in some states

than in others. We do, however, need that the observer always prefers for trade to occur: $w(b,s) \geq 0$, for all (b,s) in the support of μ.

We then say that a value x for the observer's criterion is a *robust prediction* if, for every information structure $(\mathcal{I}_B, \mathcal{I}_S, \nu)$, there exists an equilibrium (σ_B, σ_S) satisfying $W(\sigma_B, \sigma_S) \geq x$.[1] It is immediate that there is some maximum robust prediction. We wish to characterize what this value is.

Our analysis will also lead us naturally to look at symmetric information structures, where both agents have the same information. Explicitly, we say the information structure is *symmetric* if $\mathcal{I}_B = \mathcal{I}_S$ and $\eta_B = \eta_S$ with probability 1 under ν. We say that a value x is a *robust prediction with symmetric information* if, for every symmetric information structure $(\mathcal{I}_B, \mathcal{I}_S, \nu)$, there exists an equilibrium (σ_B, σ_S) satisfying $W(\sigma_B, \sigma_S) \geq x$.

3 Results

3.1 Event decomposition

Let's jump to the punch line. To identify how bad an equilibrium outcome is (from the observer's point of view), it suffices to describe when the agents fail to trade. Our two main results describe these possible no-trade events. The first main result says that for any information structure, there exists an equilibrium in which the event of no trade is at most the union of two other events, one on which the buyer has a negative expected gain from trading at price p, and one on which the seller has a negative expected gain from trading. (It does *not* say, however, that the buyer declines trade on the first sub-event, and the seller declines on the second.) The second main result is a sort of converse: for any event that has such a decomposition into two sub-events, there is an information structure under which no trade can occur there. Thus, together, these two results characterize the maximal possible no-trade events. Moreover, the second result states that one can choose the information structure to be symmetric — that is, both players have identical information. We will give the results, and then, before proceeding to the proofs (Subsection 3.3), will first lay out how they can be used to compute the maximum robust prediction (Subsection 3.2).

To state the results succinctly, we first need a little more notation. Consider any

[1] The term "prediction" is a bit of a misnomer since, as already pointed out, it depends on equilibrium selection assumptions. "Attainable value" might be more descriptive, but we keep "prediction" for simplicity.

given information structure $(\mathcal{I}_B, \mathcal{I}_S, \nu)$ and any given equilibrium (σ_B, σ_S). Let $\epsilon_B \sim U[0,1]$ be the private random variable which the buyer uses to implement any mixing that his strategy calls for — say, the buyer agrees to trade when $\epsilon_B \leq \sigma_B(\eta_B)$ — and similarly $\epsilon_S \sim U[0,1]$ for the seller. So outcomes are defined over the probability space $\Omega = \mathbb{R}^2 \times \mathcal{I}_B \times \mathcal{I}_S \times [0,1]^2$, with the probability measure $\widetilde{\nu}$ given by the product of ν with the uniform measure on $[0,1]^2$. The event of no trade, $NT \subseteq \Omega$, consists of those realizations of $(b, s, \eta_B, \eta_S, \epsilon_B, \epsilon_S)$ for which $\epsilon_B > \sigma_B(\eta_B)$ or $\epsilon_S > \sigma_S(\eta_S)$.

Proposition 3.1. *Let $(\mathcal{I}_B, \mathcal{I}_S, \nu)$ be any information structure. There exists an equilibrium (σ_B, σ_S), and two disjoint events NT_B, NT_S defined on the corresponding probability space Ω, such that*

(i) $\int_{NT_B} (b - p)\, d\widetilde{\nu} < 0$;

(ii) $\int_{NT_S} (p - s)\, d\widetilde{\nu} < 0$;

(iii) $NT \subseteq NT_B \cup NT_S$.

For the converse proposition, we let $L = \{T, B, S\}$ be a set of labels. The labels B and S will represent the events NT_B and NT_S; the label T represents the complementary event, on which trade occurs. The converse says that given any construction of the events NT_B and NT_S — which we represent by a joint distribution of (b, s) and the label $l \in L$ — such that the buyer's expected gain on NT_B and seller's on NT_S are both negative, there is some information structure for which $NT_B \cup NT_S$ is indeed a no-trade event in any equilibrium.

Proposition 3.2. *Let $\widetilde{\mu}$ be any probability distribution on $\mathbb{R}^2 \times L$, with marginal μ on $\mathbb{R}^2$. Suppose that $\int_{l=B} (b - p)\, d\widetilde{\mu} < 0$ and $\int_{l=S} (p - s)\, d\widetilde{\mu} < 0$. Then there exists a pair of finite sets $\mathcal{I}_B$ and $\mathcal{I}_S$, and a joint distribution ξ over $\mathbb{R}^2 \times \mathcal{I}_B \times \mathcal{I}_S \times L$, such that*

(i) the marginal on $\mathbb{R}^2 \times L$ is $\widetilde{\mu}$;

(ii) for any equilibrium of the information structure given by the marginal of ξ on $\mathbb{R}^2 \times \mathcal{I}_B \times \mathcal{I}_S$, the event

$$l \in \{B, S\} \text{ and } \sigma_B(\eta_B)\sigma_S(\eta_S) > 0$$

has probability 0.

Moreover, this existence holds even if we require the information structure to be symmetric.

Our next task is to show how the results can be used to compute the maximum robust prediction for the observer's criterion, given the prior μ. In the process, we reach the observation, already mentioned in the introduction, that the maximum robust prediction is the same as the maximum robust prediction with symmetric information. This observation emerges from the first step in the computation, recorded below:

Corollary 3.3. *The following are equivalent, for a real number x:*

(a) x is a robust prediction;

(b) x is a robust prediction with symmetric information;

(c) $x \leq \int_{\mathbb{R}^2} w(b,s)\, d\mu - \sup_{\widetilde{\mu}} \int_{l\in\{B,S\}} w(b,s)\, d\widetilde{\mu}$, where the supremum is over all distributions $\widetilde{\mu}$ on $\mathbb{R}^2 \times L$, having marginal μ on $\mathbb{R}^2$ and satisfying $\int_{l=B}(b-p)\, d\widetilde{\mu} < 0$ and $\int_{l=S}(p-s)\, d\widetilde{\mu} < 0$.

Proof: Clearly (a) implies (b): if a prediction of x is valid for any arbitrary information structure, it is valid for any symmetric information structure.

For (b) implies (c), suppose that $x > \int_{\mathbb{R}^2} w(b,s)\, d\mu - \sup_{\widetilde{\mu}} \int_{l\in\{B,S\}} w(b,s)\, d\widetilde{\mu}$. So there exists some $\widetilde{\mu}$ as in Proposition 3.2 such that $x > \int_{l=T} w(b,s)\, d\widetilde{\mu}$. Let ξ be given by that proposition, and $(\mathcal{I}_B, \mathcal{I}_S, \nu)$ the corresponding symmetric information structure; then for any equilibrium (σ_B, σ_S),

$$\begin{aligned} W(\sigma_B,\sigma_S) &= \int \sigma_B(\eta_B)\sigma_S(\eta_S) w(b,s)\, d\xi \\ &= \int_{l=L} \sigma_B(\eta_B)\sigma_S(\eta_S) w(b,s)\, d\xi \\ &\leq \int_{l=L} w(b,s)\, d\xi \\ &< x \end{aligned}$$

with the second line coming from conclusion (ii) of Proposition 3.2. Thus, x is not a robust prediction with symmetric information.

For (c) implies (a), suppose x satisfies the given condition. Let $(\mathcal{I}_B, \mathcal{I}_S, \nu)$ be any information structure. Let (σ_B, σ_S) and Ω, NT_B, NT_S be as given by Proposition 3.1. Let $l \in L$ be a random variable on Ω, with value B on NT_B, S on NT_S, and T otherwise; and

put $\widetilde{\mu}$ for the marginal distribution of (b, s, l). Then

$$\begin{aligned} W(\sigma_B, \sigma_S) &= \int_{\Omega \setminus NT} w(b,s)\, d\widetilde{\nu} \\ &\geq \int_{\Omega \setminus (NT_B \cup NT_S)} w(b,s)\, d\widetilde{\nu} \\ &= \int_{\mathbb{R}^2} w(b,s)\, d\mu - \int_{l \in \{B,S\}} w(b,s)\, d\widetilde{\mu} \\ &\geq x. \end{aligned}$$

Hence, x is a robust prediction. □

3.2 Maximal no-trade events

It remains, then, to calculate the supremum in (c) of Corollary 3.3 — essentially, the worst possible total weight (by the observer's criterion) of no-trade events.

We first intuitively describe the worst possible no-trade events in the benchmark case where the observer is concerned with expected gains from trade, $w(b, s) = b - s$. In this case, imagine that all possible value realizations are sorted by the ratio of gains from trade that accrue to the buyer, $(b - p)/(b - s)$. The buyer's no-trade event NT_B can be formed by taking all the realizations with ratios below a cutoff, where the cutoff is determined by the condition that the buyer's expected value on this event should equal the price p. Similarly, the seller's no-trade event NT_S consists of all realizations with high ratios, where the cutoff is determined by the seller's expected value equaling p. These events are illustrated in Figure 1: the gray heat map represents the density of the prior distribution of (b, s); the horizontally hatched region is NT_B and the diagonally hatched region is NT_S. If the events were to overlap, then the worst-case prediction would be zero trade.

For more general criteria w, it will still hold that the worst-case no-trade events will have NT_B consist entirely of lower buyer-gains ratios than NT_S. But in general, NT_B will consist of those realizations for which $(b - p)/w$ is as low as possible (subject to the buyer-gains ratio being below some cutoff), and NT_S will consist of those realizations for which $(p - s)/w$ is as low as possible (subject to the buyer-gains ratio being above its cutoff).

The rest of this subsection will formalize these ideas in full pedantic detail. We first show that the worst-case no-trade events NT_B and NT_S can be separated in terms of

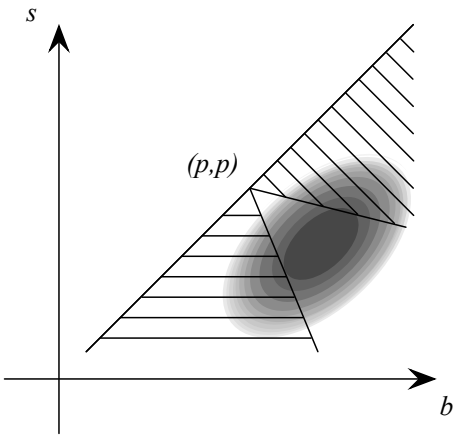

Figure 1: Worst-case no-trade events (criterion = gains from trade)

their buyer-gains ratios. Explicitly, we show that there exists $\alpha \in (0, \infty)$ such that NT_B consists only of realizations with $(b-p) \leq \alpha(p-s)$, and NT_S consists only of realizations with $(b-p) \geq \alpha(p-s)$. Moreover, if equality holds for a positive mass of realizations, then all such realizations should be split in the same proportion between NT_B and NT_S — we call this proportion β.

We will express this formally in the language of measures rather than events: The supremum in Corollary 3.3(c) equals the supremum $\int_{\mathbb{R}^2} w(b,s)d(\mu_B + \mu_S)$ over all pairs of measures μ_B, μ_S such that $\int_{\mathbb{R}^2}(b-p)\,d\mu_B < 0$, $\int_{\mathbb{R}^2}(p-s)\,d\mu_S < 0$, and $\mu_B + \mu_S \leq \mu$. Call such a pair of measures *valid*.

To state the separation lemma explicitly, given $\alpha, \beta \in [0,1]$, we define the events

$$\begin{aligned} E^\alpha_< &= \{(b,s) \mid (b-p) < \alpha(b-s)\}, \\ E^\alpha_= &= \{(b,s) \mid (b-p) = \alpha(b-s)\}, \\ E^\alpha_> &= \{(b,s) \mid (b-p) > \alpha(b-s)\}, \end{aligned}$$

and define two measures $\overline{\mu}^{\alpha,\beta}_B, \overline{\mu}^{\alpha,\beta}_S$ by

$$\overline{\mu}^{\alpha,\beta}_B(E) = \mu(E \cap E^\alpha_<) + \beta\mu(E \cap E^\alpha_=), \qquad \overline{\mu}^{\alpha,\beta}_S(E) = \mu(E \cap E^\alpha_>) + (1-\beta)\mu(E \cap E^\alpha_=) \quad (3.1)$$

for any event E. Note that $\overline{\mu}^{\alpha,\beta}_B + \overline{\mu}^{\alpha,\beta}_S = \mu$. (We will often write these without the superscript α, β.)

We will then say that a pair of measures (μ_B, μ_S) is (α, β)-*separated* if $\mu_B \leq \overline{\mu}^{\alpha,\beta}_B$ and

$\mu_S \le \overline{\mu}_S^{\alpha,\beta}$.

Lemma 3.4. *Let (μ_B, μ_S) be a valid pair. Then there exists a valid pair $(\widehat{\mu}_B, \widehat{\mu}_S)$ that is (α, β)-separated, for some α, β, and such that $\widehat{\mu}_B + \widehat{\mu}_S = \mu_B + \mu_S$.*

The proof is straightforward: With (μ_B, μ_S) given, any choice of parameters α, β specifies a way of redividing the mass $\mu_B + \mu_S$ into $\widehat{\mu}_B$ and $\widehat{\mu}_S$. There is some range of pairs (α, β) for which the needed inequality $\widehat{\mu}_B \le \overline{\mu}_B^{\alpha,\beta}$ is satisfied, and a corresponding range for $\widehat{\mu}_S$; we just need to show that these two parameter ranges overlap. The details are in Appendix A.

Lemma 3.4 shows that in our search for the supremum of $\int w(b,s)\, d(\mu_B + \mu_S)$ over valid pairs of measures, we can restrict ourselves to looking at valid pairs that are (α, β)-separated for some α, β.

So, for any given α, β, define $Y(\alpha, \beta)$ to be the supremum of $\int w(b,s)\, d(\mu_B + \mu_S)$ over valid pairs that are (α, β)-separated. We just need a way to compute $Y(\alpha, \beta)$ for given α and β, and then in a subsequent round we optimize over α, β.

It is evident that

$$Y(\alpha, \beta) = \sup_{\mu_B} \int w(b,s)\, d\mu_B + \sup_{\mu_S} \int w(b,s)\, d\mu_S,$$

where the first supremum is over all measures $\mu_B \le \overline{\mu}_B$ satisfying $\int (b-p)\, d\mu_B < 0$, and the second is over all measures $\mu_S \le \overline{\mu}_S$ satisfying $\int (p-s)\, d\mu_S < 0$. We denote these two separate suprema by $Y_B(\alpha, \beta), Y_S(\alpha, \beta)$.

These separate suprema can be calculated by the greedy algorithm that takes mass that (for μ_B) minimizes the ratio $(b-p)/w$, up until the point where the total integral of $b - p$ is zero; or (for μ_S) minimizes $(p-s)/w$, up until the integral of $p - s$ is zero.

Let us give a precise statement. For $\gamma > 0$ and $\delta \in [0, 1]$, define

$$F_<^\gamma = \{(b,s) \mid b - p < \gamma w(b,s)\}, \qquad F_=^\gamma = \{(b,s) \mid b - p = \gamma w(b,s)\}.$$

Then $F_<^\gamma$ is increasing in γ, and so $\int_{F_<^\gamma} (b-p)\, d\overline{\mu}_B$ is also increasing in γ, since the pairs that are in $F_<^\gamma$ but not in $F_<^{\gamma'}$ for $\gamma' < \gamma$ must satisfy $b - p > 0$. Let $\gamma_B^* \in (0, \infty]$ be the supremum of values such that $\int_{F_<^\gamma} (b-p)\, d\overline{\mu}_B < 0$. (This integral is negative at $\gamma = 0$.) If $\gamma_B^* < \infty$ then $\int_{F_<^{\gamma_B^*}} (b-p)\, d\overline{\mu}_B + \delta \int_{F_=^{\gamma_B^*}} (b-p)\, d\overline{\mu}_B$ is weakly increasing in $\delta \in [0, 1]$, and is nonnegative at $\delta = 1$; let δ_B^* be the supremum of values for which it is < 0. The expression must be equal to 0 at $\delta = \delta_B^*$.

For the seller's no-trade event, we perform a completely analogous computation, substituting $p - s$ for $b - p$ and $\overline{\mu}_S$ for $\overline{\mu}_B$, and defining events

$$G^\gamma_< = \{(b,s) \mid p - s < \gamma w(b,s)\}, \qquad G^\gamma_= = \{(b,s) \mid p - s = \gamma w(b,s)\}.$$

This gives values γ^*_S and δ^*_S.

Lemma 3.5. *If $\gamma^*_B = \infty$ then $Y_B(\alpha,\beta) = \int_{\mathbb{R}^2} w(b,s)\, d\overline{\mu}_B$. Otherwise,*

$$Y_B(\alpha,\beta) = \int_{F^{\gamma^*_B}_<} w(b,s)\, d\overline{\mu}_B + \delta^*_B \int_{F^{\gamma^*_B}_=} w(b,s)\, d\overline{\mu}_B.$$

Similarly, if $\gamma^*_S = \infty$ then $Y_S(\alpha,\beta) = \int_{\mathbb{R}^2} w(b,s)\, d\overline{\mu}_S$, and otherwise

$$Y_S(\alpha,\beta) = \int_{G^{\gamma^*_S}_<} w(b,s)\, d\overline{\mu}_S + \delta^*_S \int_{G^{\gamma^*_S}_=} w(b,s)\, d\overline{\mu}_S.$$

The proof is in Appendix A.

Finally, we can summarize our work in the following procedure to compute the observer's maximum robust prediction, given the prior distribution μ.

1. For each choice of $\alpha, \beta \in [0,1]$, split μ into $\overline{\mu}_B$ and $\overline{\mu}_S$ by (3.1).
2. Use the greedy algorithm on this $\overline{\mu}_B$ and $\overline{\mu}_S$ — taking the mass with the lowest ratio $(b-p)/w$ and $(p-s)/w$, respectively — to compute $Y_B(\alpha,\beta)$ and $Y_S(\alpha,\beta)$, as described in Lemma 3.5. This determines $Y(\alpha,\beta) = Y_B(\alpha,\beta) + Y_S(\alpha,\beta)$, the worst possible total weight of no-trade events subject to the given values of α and β.
3. Finally, as given by Corollary 3.3, the maximum robust prediction equals $\int_{\mathbb{R}^2} w(b,s)\, d\mu - \sup_{\alpha,\beta} Y(\alpha,\beta)$.

We note that the brief description given earlier for the benchmark case $w(b,s) = b - s$ — where the no-trade event NT_B is formed by taking the realizations with the lowest ratio $(b-p)/(b-s)$, and NT_S is formed by taking the realizations with the highest ratio — immediately follows as a special case.

3.3 Proofs of main results

We now give the proofs of the main results, which have so far been cruelly withheld.

The proof of Proposition 3.1 — existence of a "good" equilibrium for any information structure — is nonconstructive. We consider a sequence of constrained games, where the players are not always permitted to choose freely to accept or reject trade; instead, for some realizations of the signals, we force them to accept trade. Initially, we force both players to trade for all realizations of their signals. We then gradually unconstrain the signal realizations, one by one, and apply the Nash existence theorem to each constrained game. As long as the equilibrium of the constrained game is not also an equilibrium of the unconstrained game, we can express one player's desire to deviate as an inequality; combining these inequalities leads to the desired event decomposition.

Proof of Proposition 3.1: We successively define sequences of signal sets $\mathcal{J}_B^k \subseteq \mathcal{I}_B, \mathcal{J}_S^k \subseteq \mathcal{I}_S$ and functions $\lambda_B^k, \lambda_S^k : \mathcal{I}_B \times \mathcal{I}_S \to [0,1]$, for $k = 0, 1, \ldots$. These sets and functions will be made to satisfy the following conditions:

(a) $\lambda_B^k(\eta_B, \eta_S) = 0$ whenever $\eta_B \in \mathcal{J}_B^k$;

(b) $\lambda_S^k(\eta_B, \eta_S) = 0$ whenever $\eta_S \in \mathcal{J}_S^k$;

(c) if $(\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k$, then $\lambda_B^k(\eta_B, \eta_S) + \lambda_S^k(\eta_B, \eta_S) \geq 1$;

(d) if $\mathcal{J}_B^k \neq \mathcal{I}_B^k$, then $\int \lambda_B^k(\eta_B, \eta_S) \cdot (b - p)\, d\nu < 0$;

(e) if $\mathcal{J}_S^k \neq \mathcal{I}_S^k$, then $\int \lambda_S^k(\eta_B, \eta_S) \cdot (p - s)\, d\nu < 0$.

$\mathcal{J}_B^k$ will be the set of signal realizations in which the buyer is forced to accept trade in the kth constrained game; similarly for the seller and $\mathcal{J}_S^k$. λ_B^k and λ_S^k will be weights derived from the deviation inequalities along the way.

For the base case, we take $\mathcal{J}_B^0 = \mathcal{I}_B$, $\mathcal{J}_S^0 = \mathcal{I}_S$, and λ_B^0, λ_S^0 identically zero. It is clear that (a) and (b) hold, and (c)–(e) are vacuous.

Now suppose these sets and functions have been defined for some k. Consider the Bayesian game where each player learns his signal according to ν, and agrees or declines to trade, with the constraint that the buyer must agree to trade whenever $\eta_B \in \mathcal{J}_B^k$, and likewise the seller must agree whenever $\eta_S \in \mathcal{J}_S^k$. That is, the (mixed) strategy space of the buyer is the set of $\sigma_B : \mathcal{I}_B \to [0,1]$ such that $\sigma_B(\eta_B) = 1$ whenever $\eta_B \in \mathcal{J}_B^k$, and likewise for the seller; and the payoffs are given by (2.1). This game has a Bayesian Nash equilibrium, call it (σ_B, σ_S).

Suppose that (σ_B, σ_S) is not an equilibrium of the original, unconstrained game. In this case we will define $\mathcal{J}_B^{k+1}, \mathcal{J}_S^{k+1}, \lambda_B^{k+1}, \lambda_S^{k+1}$. One of the players has a profitable deviation, say the buyer (the argument if it is the seller is totally analogous). In particular, there

is at least one signal η_B^* on which the buyer benefits from deviating. That is, there is σ_B' that agrees with σ_B for all signals except η_B^*, and such that

$$u_B(\sigma_B', \sigma_S) > u_B(\sigma_B, \sigma_S). \tag{3.2}$$

We must have $\eta_B^* \in \mathcal{J}_B^k$, because otherwise the deviation σ_B' would be allowed in the constrained game, and (3.2) contradicts the assumption that (σ_B, σ_S) was an equilibrium of the constrained game. Therefore, $\sigma_B(\eta_B^*) = 1$, and $\sigma_B'(\eta_B^*) < 1$. So (3.2) implies

$$\int_{\eta_B = \eta_B^*} \sigma_S(\eta_S)(b-p)\, d\nu < 0. \tag{3.3}$$

Define $\mathcal{J}_B^{k+1} = \mathcal{J}_B^k \setminus \{\eta_B^*\}$, and define

$$\lambda_B^{k+1}(\eta_B, \eta_S) = \begin{cases} \sigma_S(\eta_S) & \text{if } \eta_B = \eta_B^*, \\ \lambda_B^k(\eta_B, \eta_S) & \text{otherwise.} \end{cases}$$

Also define $\mathcal{J}_S^{k+1} = \mathcal{J}_S^k$ and $\lambda_S^{k+1} = \lambda_S^k$.

We check that (a)-(e) are satisfied for step $k+1$. It is straightforward to see that (a) for $k+1$ follows from (a) for k. For (c), we only need to check the cases where $\eta_B = \eta_B^*$. There are two possibilities. If $\eta_S \notin \mathcal{J}_S^k$, then

$$\begin{aligned} \lambda_B^{k+1}(\eta_B, \eta_S) + \lambda_S^{k+1}(\eta_B, \eta_S) &\geq \lambda_B^k(\eta_B, \eta_S) + \lambda_S^{k+1}(\eta_B, \eta_S) \\ &= \lambda_B^k(\eta_B, \eta_S) + \lambda_S^k(\eta_B, \eta_S) \\ &\geq 1. \end{aligned}$$

Here the first line is because $\lambda_B^k(\eta_B, \eta_S) = 0$ (by (a) for k); the second is because $\lambda_S^{k+1} = \lambda_S^k$; the third is by (c) for k. If on the other hand $\eta_S \in \mathcal{J}_S^k$, then $\lambda_B^k(\eta_B, \eta_S) = \sigma_S(\eta_S) = 1$ already. So (c) holds. For (d), we already know $\int \lambda_B^k(\eta_B, \eta_S)(b-p)\, d\nu \leq 0$. And

$$\begin{aligned} &\int \lambda_B^{k+1}(\eta_B, \eta_S)(b-p)\, d\nu - \int \lambda_B^k(\eta_B, \eta_S)(b-p)\, d\nu \\ &\qquad = \int_{\eta_B = \eta_B^*} (\lambda_B^{k+1}(\eta_B, \eta_S) - \lambda_B^k(\eta_B, \eta_S))(b-p)\, d\nu \\ &\qquad = \int_{\eta_B = \eta_B^*} \sigma_S(\eta_S)(b-p)\, d\nu \\ &\qquad < 0 \end{aligned}$$

by (3.3). Finally, (b) and (e) hold since $\mathcal{J}_S^{k+1} = \mathcal{J}_S^k$ and $\lambda_S^{k+1} = \lambda_S^k$.

Now, at each step k of this construction, the sets $\mathcal{J}_B^k, \mathcal{J}_S^k$ become weakly smaller, and one of them becomes strictly smaller. By finiteness, the process must stop at some k. This can only happen when the constrained equilibrium (σ_B, σ_S) is an equilibrium of the unconstrained game. This will be the equilibrium claimed in the proposition. It remains to define events NT_B, NT_S.

First, we can change λ_B^k and λ_S^k if necessary so that the inequality in condition (c) becomes an equality. To see this, consider any $(\eta_B^*, \eta_S^*) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k$. At least one of

$$\int_{(\eta_B,\eta_S)=(\eta_B^*,\eta_S^*)} (b-p)\, d\nu, \qquad \int_{(\eta_B,\eta_S)=(\eta_B^*,\eta_S^*)} (p-s)\, d\nu$$

is nonnegative, since their sum is nonnegative. If the former, we can replace $\lambda_B^k(\eta_B^*, \eta_S^*)$ by the lower value $1 - \lambda_S^k(\eta_B^*, \eta_S^*)$ (keeping all other values of λ_B^k the same); this will make (c) hold with equality at this pair and will preserve (d) since the left side of the inequality there becomes weakly smaller. Likewise, in the latter case we replace $\lambda_S^k(\eta_B^*, \eta_S^*)$ by $1 - \lambda_B^k(\eta_B^*, \eta_S^*)$. Doing this for each signal pair, we ensure that (c) is an equality for each signal pair where it applies.

Now suppose momentarily that $\mathcal{J}_B^k \neq \mathcal{I}_B$ and $\mathcal{J}_S^k \neq \mathcal{I}_S$. Let

$$\begin{aligned} NT_B \quad &\text{be the event} \quad ((\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k \text{ and } \epsilon_B \leq \lambda_B^k(\eta_B, \eta_S)); \\ NT_S \quad &\text{be the event} \quad ((\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k \text{ and } \epsilon_B > \lambda_B^k(\eta_B, \eta_S)). \end{aligned}$$

We check that conditions (i)-(iii) in the proposition statement hold. For (i), conditional on any realizations of (b, s, η_B, η_S), the probability of NT_B is $\lambda_B^k(\eta_B, \eta_S)$ (this is true also when $(\eta_B, \eta_S) \in \mathcal{J}_B \times \mathcal{J}_S$ since then $\lambda_B^k(\eta_B, \eta_S) = 0$). So

$$\int_{NT_B} (b-p)\, d\widetilde{\nu} = \int \lambda_B^k(\eta_B, \eta_S)(b-p)\, d\widetilde{\nu} < 0.$$

Likewise for (ii), conditional on (b, s, η_B, η_S), the probability of NT_S is $\lambda_S^k(\eta_B, \eta_S)$ — this holds for $(\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k$ by equality in (c), and otherwise both sides are zero. Hence

$$\int_{NT_S} (p-s)\, d\widetilde{\nu} = \int \lambda_S^k(\eta_B, \eta_S)(p-s)\, d\widetilde{\nu} < 0.$$

And for (iii), $NT_B \cup NT_S$ is the event $(\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k$. But NT only occurs when $\sigma_B(\eta_B) < 1$ or $\sigma_S(\eta_S) < 1$, which (by construction of the constrained game) requires

$(\eta_B, \eta_S) \notin \mathcal{J}_B^k \times \mathcal{J}_S^k$.

Finally, if $\mathcal{J}_B^k$ is all of $\mathcal{I}_B$ or $\mathcal{J}_S^k$ is all of $\mathcal{I}_S$, then we use the same construction except that we will have equality in (i) or (ii), respectively. We can use small adjustments to turn these into strict inequalities: If only (i) is an equality, we can consider a small amount of probability mass on which $b - p < 0$, and add it to NT_B (if this mass already belongs to NT_S then remove it from NT_S in order to preserve disjointness); this makes (i) hold strictly and preserves (ii) and (iii). If only (ii) is an equality, proceed similarly. If both are equalities, then NT_B and NT_S as defined above are empty; we can obtain strict inequality by instead letting NT_B be any positive-probability sub-event of $b - p < 0$ and NT_S be any (disjoint) sub-event of $p - s < 0$. □

In contrast to the preceding proof, Proposition 3.2 — existence of an information structure forcing no trade on certain events — will be proved by a very simple construction: Both players simply observe the label l as their signal.

Proof of Proposition 3.2: Let $\mathcal{I}_B = \mathcal{I}_S = L$. Let (b, s, l) be distributed according to $\widetilde{\mu}$, and put $\eta_B = \eta_S = l$. This generates a distribution ξ, and the information structure is symmetric. Clearly (i) is satisfied. For (ii), we need to show that any equilibrium (σ_B, σ_S) satisfies $\sigma_B(B)\sigma_S(B) = \sigma_B(S)\sigma_S(S) = 0$. If $\sigma_S(B) > 0$, then the buyer will not want to trade when he receives signal B, since his gains from trading are $\int_{l=B}(\sigma_S(B))(b-p)\, d\widetilde{\mu} < 0$. Thus $\sigma_B(B)\sigma_S(B) = 0$. The proof that $\sigma_B(S)\sigma_S(S) = 0$ is analogous. □

3.4 Comments on sign criteria

Now that we've got those proofs behind us, let's briefly discuss the consequences of relaxing some of the assumptions on signs, as was promised some while ago.

Both parties uncertain about gains. We have assumed that the events $b - p < 0$ and $p - s < 0$ both have positive probability. What happens when one has probability zero? If, say, $p - s \geq 0$ for certain, then the seller is always willing to accept trade. On any information structure, constraining the seller to always accept, and having the buyer choose a best response, gives an equilibrium. So Proposition 3.1 becomes simpler: there exists an event NT_B over which the integral of $b - p$ is negative, and $NT = NT_B$. The converse, analogous to Proposition 3.2, says now that for any event on which the integral of $b - p$ is negative, there is a (symmetric) information structure for which trade cannot occur there. Hence, to compute the maximum robust prediction, we just need to compute the supremum of $\int w(b, s)\, d\mu_B$ over measures with $\int (b - p)\, d\mu_B < 0$ — which we do by the greedy algorithm — and then subtract it from $\int w(b, s)\, d\mu$.

Of course, if both $b - p \geq 0$ and $p - s \geq 0$ for certain, then it is always an equilibrium for both agents always to trade.

Observer prefers trade. We have required the observer's criterion w to be nonnegative. What if w could be negative — for example, the observer is concerned with the buyer's (best) expected payoff in equilibrium? Then Corollary 3.3 no longer determines exactly the maximum robust prediction, because of a gap between Propositions 3.1 and 3.2. Proposition 3.1 says that the event of no trade is *contained in* the union of two events, NT_B (where the integral of $b-p$ is negative) and NT_S (where the integral of $p-s$ is negative). If w is nonnegative, then the worst case is to have the no-trade event be all of NT_B and NT_S, and Proposition 3.2 says this can indeed happen for some information structure. However if w can have negative values on part of NT_B and NT_S, then the worst case may be even worse, having trade occur only on the parts of NT_B and NT_S where w takes on negative values.

It is still possible to use Proposition 3.1 to give a nontrivial robust prediction for the observer's payoff, in many cases, since trade must always occur everywhere outside of $NT_B \cup NT_S$, but this robust prediction may not be best possible.

Aggregate gains from trade. We have also assumed common knowledge of gains from trade — $b-s \geq 0$ for sure. Nothing changes as long as we have the weaker assumption that $\max\{b-p, p-s\} \geq 0$ for sure (and continue to require $w(b,s) \geq 0$ everywhere, which may require the observer's criterion to be something other than gains from trade).

However, if it is possible that both $b-p$ and $p-s$ are negative, then we can no longer ensure disjointness of the events NT_B, NT_S in Proposition 3.1. This is because condition (c) in the proof may be satisfied with strict inequality, and unlike before, we can no longer decrease one of the λ's to make it become an equality. This again gives us a gap between Propositions 3.1 and 3.2, since in the latter the events $l = B$ and $l = S$ are clearly disjoint. So again, Proposition 3.1 may give us a nontrivial robust prediction, but Proposition 3.2 no longer ensures that this prediction is optimal.

4 Examples

Here we give an example showing an application of the results of Section 3, as well as a couple of examples exploring some interpretive issues.

4.1 Computing the Maximum Robust Prediction

We give a simple (perhaps too simple) application of our results, showing how to compute the maximum robust prediction, in an example adapted from Morris and Shin [14]. It is common knowledge that the the good is worth $2c$ more to the buyer than it is to the seller. Most likely, it is worth $p + c$ to the buyer and $p - c$ to the seller. However, there is a small probability δ that the good is a lemon, with low value to both parties, and probability δ that it is a peach, with high value to both parties. Specifically, the common prior distribution μ is that

$$(b, s) = \begin{cases} (p - M + c, p - M - c) & \text{with probability } \delta, \\ (p + c, p - c) & \text{with probability } 1 - 2\delta, \\ (p + M + c, p + M - c) & \text{with probability } \delta. \end{cases}$$

(Here $M > c > 0$.) We take $w(b, s) = 1$ everywhere, so we are interested in robustly predicting the probability of trade; predicted gains from trade are just $2c$ times this probability. Note that the numerical illustration at the beginning of the introduction is an instance of this setup.

Since the criterion $w(b, s) = 1$ and the criterion $w(b, s) = 2c$ are equivalent in this example, the shortcut at the beginning of Subsection 3.2 applies: we form the no-trade event NT_B by carving out probability mass with the lowest possible buyer-gains ratios, up until the point where the buyer's conditional expected value equals the price p; and we form the no-trade event NT_S by carving out probability mass with the highest possible buyer-gains ratios, until the seller's expected value equals p. If these two events end up overlapping, then the best robust prediction is zero trade.

More precisely, there are two cases depending on parameters:

- If $\delta M/c \leq 1/2$, then the maximal amount of probability mass that can belong to NT_B is $\delta M/c$ — consisting of the δ probability of lemon realizations, together with a $\delta(M - c)/c$ probability mass of normal realizations. Likewise the maximal NT_S consists of the δ probability of peach realizations and a $\delta(M - c)/c$ probability mass of normal realizations. Therefore, by Corollary 3.3, the maximum robust prediction is $1 - 2\delta M/c$. That is, for any information structure, there is an equilibrium where trade occurs with probability at least $1 - 2\delta M/c$; and this bound is sharp, even with the restriction to symmetric information.

  To be even more explicit, we describe an information structure approaching the

bound: Both parties receive the same signal, $\eta_B = \eta_S = \eta \in \{T, B, S\}$. The joint distribution of values and signals is as shown in Table 4.1. (Note that the formatting of this table is different from Table 1; here rows are values and columns are signals.) Here $\epsilon > 0$ is arbitrarily small. Thus, under the signal B — which is a noisy signal of the lemon state — trade cannot occur because the buyer's expected value is less than p. Under the peach signal S, trade cannot occur because the seller's expected value is greater than p. So trade occurs with probability at most $1 - 2\delta M/c + 2\epsilon$.

Values	$\eta = B$	$\eta = T$	$\eta = S$
$(p - M + c, p - M - c)$	δ	0	0
$(p + c, p - c)$	$\delta\frac{M-c}{c} - \epsilon$	$1 - 2\delta\frac{M}{c} + 2\epsilon$	$\delta\frac{M-c}{c} - \epsilon$
$(p + M + c, p + M - c)$	0	0	δ

Table 2: Distribution of values and (symmetric) signals

- If $\delta M/c > 1/2$, then the best possible robust prediction is 0: the information may be structured so that no trade can occur in equilibrium.

  One possible information structure that yields no trade (not the only one) is to have a shared signal $\eta \in \{B, S\}$, jointly distributed with the values as shown in Table 4.1. Under signal B, the buyer's expected value is less than p; under signal S, the seller's expected value is more than p.

Values	$\eta = B$	$\eta = S$
$(p - M + c, p - M - c)$	δ	0
$(p + c, p - c)$	$\frac{1}{2} - \delta$	$\frac{1}{2} - \delta$
$(p + M + c, p + M - c)$	0	δ

Table 3: Distribution of values and (symmetric) signals

4.2 No-Trade Equilibria

As mentioned in the introduction, the interpretation of our results as a positive prediction about the amount of trade depends on an implicit assumption about equilibrium selection, since there is also an equilibrium in which neither agent ever accepts trade, for any information structure. In some cases, one can brush aside such a bad equilibrium using a

standard refinement such as elimination of weakly dominated strategies, or more generally trembling-hand perfection. For example, under any *symmetric* information structure, undominated strategies will imply that each agent agrees to trade when his expected payoff from successfully trading is strictly positive; hence the prediction from Corollary 3.3 applies to any equilibrium in undominated strategies, not just the best equilibrium.

However, this kind of refinement does not always get around the issue. We now show an example, building on the previous subsection, in which there can be trembling-hand perfect equilibria with no trade, even though the good equilibrium outcome involves trade most of the time.

Let μ be as given in Subsection 4.1, for some parameter values with $\delta M/c$ small, so that for any information structure there is an equilibrium with a high probability of trade. Now consider the following information structure. The signal sets are $\mathcal{I}_B = \mathcal{I}_S = \{L, N, P\}$. The letters stand for "lemon, normal, peach," and the first and last signals are perfectly informative while the middle signal is imperfectly informative. Specifically, conditional on the realized values (b, s), both players' signals are independently drawn from the same distribution, which is given by Table 4.2.

Values	$Pr(L)$	$Pr(N)$	$Pr(P)$
$(p - M + c, p - M - c)$	1/2	1/2	0
$(p + c, p - c)$	0	1	0
$(p + M + c, p + M - c)$	0	1/2	1/2

Table 4: Distribution of signals, conditional on values

There are some signal realizations for which the players have (weakly) dominant actions: If the buyer receives L, he knows the values are $(p - M + c, p - M - c)$ for sure, so he does not accept trade, in any trembling-hand perfect equilibrium. Similarly, if the seller receives L, he does accept. If the buyer receives P, he accepts; if the seller receives P, he does not accept.

Let (σ_B, σ_S) be the following strategy profile: the buyer accepts only when his signal is P, and the seller accepts only when his signal is L. To check that this is a trembling-hand perfect equilibrium, it suffices to check that each player is playing a strict best reply to the other's strategy when his own signal is N, since it follows that each player's strategy is a best reply to any sufficiently small tremble. Consider the buyer's strategy when his signal is N. From his point of view, any of the three value pairs — and any of the seller's signals — can occur with positive probability. But if he accepts trade, the trade will only

occur if the seller's signal is L, in which case trade is definitely bad for him. So the buyer strictly loses by agreeing to trade on signal N. Similarly for the seller.

In this equilibrium, trade only occurs if the buyer receives signal P and the seller receives L; but this can never happen.

4.3 Alternative Mechanisms

As mentioned in the introduction, one interpretation of our results is that breakdown of trade is not intrinsically due to asymmetric information. However, it is important to qualify this interpretation by pointing out that we have assumed a particular mechanism for trade, namely the posted price p, which is fixed independent of the information received by the buyer and seller. With other mechanisms, it may no longer be true that the maximum robust prediction with symmetric information is the same as the maximum robust prediction without symmetric information.

For example, consider instead a double auction mechanism: the buyer names a price p_B, and the seller names a price p_S; if $p_B < p_S$ then no trade takes place, and if $p_B \geq p_S$ then trade happens at price $(p_B + p_S)/2$. For any μ, and any symmetric information structure, there is an equilibrium in which the parties always trade: For each realization of the signal η, pick any price $p(\eta)$ lying in between the buyer's and seller's expected values conditional on η; then it is an equilibrium for both parties, after observing η, to name the price $p(\eta)$. This mechanism realizes all gains from trade.

In view of this observation, one might ask: is it possible that, no matter what the information structure is, the buyer and seller can always come up with some suitable mechanism that realizes all (or at least most) of the gains from trade? After all, we have assumed it is common knowledge that $b \geq s$, so a classic impossibility result such as Myerson and Satterthwaite [15] does not apply. However, the answer is no: One can give examples of asymmetric information structures where no mechanism guarantees efficient trade — indeed, where it may not be possible to achieve *any* trade in equilibrium, even though there is common knowledge of gains from trade.

There does not seem to be a canonical reference for this fact in the literature, so we will digress briefly to develop an example in detail. The example is actually adopted from Akerlof [1], though his original analysis only considered posted-price mechanisms; our analysis is similar in spirit to Hendren's [8] result on impossibility of trade in insurance markets, though simpler.

Let s be uniformly distributed on $[0, 1]$, and $b = \lambda s$, where $\lambda \in (1, 2)$ is a constant. This

gives the common-prior distribution μ. Now consider the information structure where the seller knows s (and so also b) perfectly, and the buyer is completely uninformed. (This admittedly does not fit perfectly into our model, since we have previously considered only finite information structures, but arbitrarily close discrete approximations should give the same qualitative conclusion.) We will sketch a proof that there is no incentive-compatible, individually rational mechanism that achieves positive gains from trade.

Suppose such a mechanism exists. Let $p(s)$ be the probability of sale, and $t(s)$ the expected payment to the seller, when his value is s. Then his net utility from the mechanism is $U_S(s) = t(s) - sp(s)$. We have the usual incentive-compatibility condition

$$t(s) - sp(s) \geq t(s') - sp(s') \text{ for all } s, s'$$

and individual rationality

$$t(s) - sp(s) \geq 0.$$

Standard revealed-preference arguments imply that $U_S(s)$ is weakly decreasing, hence differentiable almost everywhere; $p(s)$ is weakly decreasing; and the envelope argument gives $U_S'(s) = -p(s)$ from which

$$U_S(s) = U_S(1) + \int_s^1 p(\widehat{s})\, d\widehat{s}$$

and so

$$t(s) = U_S(1) + \int_s^1 p(\widehat{s})\, d\widehat{s} + sp(s).$$

Now, when the seller's value is s, the buyer's utility from participating in the mechanism is $\lambda sp(s) - t(s)$. For the buyer to be willing to participate, the expected value of this utility should be nonnegative:

$$\int_0^1 \lambda sp(s) - t(s)\, ds \geq 0.$$

Plugging in for $t(s)$, we have

$$\int_0^1 \left(\lambda sp(s) - U_S(1) - \int_s^1 p(\widehat{s})\, d\widehat{s} - sp(s) \right) ds \geq 0. \tag{4.1}$$

But by a change of variables, we have $\int_0^1 (\int_s^1 p(\widehat{s})\, d\widehat{s})\, ds = \int_0^1 sp(s)\, ds$. So (4.1) simplifies

to

$$\int_0^1 ((\lambda - 2)sp(s) - U_S(1))\ ds \geq 0.$$

The left-hand side is nonpositive. So the equality can hold only if $U_S(1) = 0$ and $p(s) = 0$ everywhere — which means that trade never occurs.

The upshot of this discussion is that our sharp prediction on attainable gains from trade is sensitive to the choice of mechanism. In important cases (e.g. symmetric information structures) an easy change to the mechanism could realize more gains from trade than our posted-price mechanism. However, it may happen that even with the best mechanism, not all gains from trade can be realized. A natural question to ask in future work is to obtain robust predictions for equilibrium gains from trade when the parties use the best mechanism, instead of imposing the posted-price mechanism as we have here (or any other) — and to identify the information structure that makes trading most difficult even when the parties can choose the mechanism. However, this problem seems substantially more difficult.

5 Words at the End

5.1 Interpretation

Now it's time to fulfill the promise in the introduction, to discuss possible economic interpretations of our main results. We stress, however, that the question of economic interpretation is basically separate from the methodological purpose of the paper, which has already been discussed.

A key assumption is that the observer knows the distribution of buyer's and seller's values for the good, but does not know the information structure and does not directly observe the trading outcomes. Thus, it makes sense to think of the observer not as an econometrician who has past trading data, but perhaps as a planner trying to orchestrate future trades, with limited foresight of the relevant environment.

For example, one might imagine that a buyer and seller are considering contracting on a specialized widget, which they may or may not actually wish to trade in the future, but which requires investment in technology today in order to be able to trade later. Our model applies if they can currently foresee the physical circumstances that affect each party's value for the widget, but cannot anticipate how informed each party will be when the time comes to trade. A lower bound for the attainable gains from trade can

potentially provide an immediate guarantee that the investment is worthwhile.[2]

A related application might be to a regulator designing a financial market, in which agents might be able to trade some security whose value depends on future events. If the regulator can anticipate how the events will affect the security's value but not the details of what information the traders will have, a lower-bound result can potentially provide assurance that there will still be enough trade in the market to warrant the fixed costs of opening the market.

A different perspective is to fit our work in with the literature on design of information structures [10, 16, 12], taking the worst-case information structure literally as a description of how an adversary might best prevent two parties from trading. This might describe, for example, a firm that tries to prevent its rival from successfully trading with a supplier by putting in place a technology that reveals to them information relevant to their trade.

Finally, one more economic interpretation of our results is as a counterpoint to the literature on how asymmetric information leads to breakdown of trade. In particular, recent work such as Morris and Shin [14] emphasizes the role of higher-order beliefs in trade breakdown. Although it may indeed sometimes happen ex post that gains from trade go unrealized for reasons traceable to higher-order beliefs, our results show that from the ex-ante perspective, higher-order beliefs are not needed to explain the breakdown of trade. That is, given the known distribution over values, the probability of trade breakdown that can be explained using higher-order beliefs is no worse than may occur with very simple and indeed symmetric information structures. This finding builds in a natural way on the earlier work of Kessler [11] and Levin [13] showing that the extent of trade breakdown in lemons markets is generally non-monotone in the amount of information asymmetry. However, an important caveat to this interpretation is that it depends on our assumption of a posted-price mechanism for trade. As discussed in Subsection 4.3 above, a different mechanism could lead to different predictions.

More generally, a few words should be said about our assumption of a posted-price mechanism and its importance. As we have seen, this assumption is limiting, both in terms of the sharpness of our characterization — we show how to find the highest possible

[2]In our model, there is common knowledge of gains from trade. In this case, having a lower-bound guarantee seems superfluous in this contracting story: the parties simply could agree up front to trade with probability 1, at a price that splits the ex-ante gains from trade. However, the model fits the following variant: The buyer will find out tomorrow whether he wants the widget (in which case gains from trade are positive) or doesn't want the widget (gains are negative), and there may be additional information as well, of unknown structure. Ex ante, the buyer is unlikely to want the widget, so that simply contracting to sell is inefficient. The model then describes what happens conditional on the buyer wanting the widget.

robust prediction for gains from trade, but this is no longer sharp if the agents are allowed to choose a different mechanism — and our observations about the nature of the worst-case information structure. One unimpressive defense that can be given is that we simply follow the literature — e.g. [1, 7, 14] — in adopting this simple trading mechanism, in order to better focus attention on the question of information structure. Another point is that our main result is a lower bound on the attainable gains from trade; it would continue to hold *a fortiori* if the parties were also allowed to use other mechanisms, instead of being restricted to a posted price. In particular, imagine a double auction mechanism as in Subsection 4.3. Any equilibrium of our posted-price mechanism can be translated into an equilibrium of the double-auction mechanism: reinterpret "accepting price p" as a bid of p in the double auction, and reinterpret "rejecting price p" as making an unacceptable bid in the double auction (a bid outside the support of values, which the other party would never want to accept). This produces the same outcome as the original equilibrium of the posted-price mechanism. Thus our sharp lower bound on attainable trade in the posted-price mechanism is also a valid lower bound for the double auction mechanism, which has the advantage of being "parameter-free," unlike the posted price mechanism which has the pesky p exogenously given.

5.2 Future directions

We wrap up by quickly surveying directions for future exploration. On the technical side, the sharp characterization of robust predictions of trade calls out to be extended to allow for $b < s$, and more generally to allow negative values of the observer's criterion w. The other major direction, already pointed out in Subsection 4.3, would be to ask about the best equilibrium outcome of the best trading mechanism, rather than a specific posted-price mechanism. More incremental extensions could keep the restriction to a very simple trading mechanism, but consider trade in multiple units of a good, or multiple goods.

From the methodological point of view, the role of this paper is to ask what predictions can be made about economic interactions without knowing the details of the information structure. We have focused here on one of the simplest possible economic transactions, exchange of a single indivisible good. Aside from generalizing our results within the pure exchange setting, it will be natural for future work to look at other similar workhorse models — production, moral hazard, coordination games, public good provision — and see where analogous approaches yield interesting results.

A Some Boring Details

Proof of Lemma 3.4: For conciseness put $\mu_+ = \mu_B + \mu_S$. As α ranges over $[0,1]$, the event $E^\alpha_<$ is increasing in α. (This depends on the fact that $b - s \geq 0$ everywhere.) Moreover, any pair (b,s) contained in one $E^\alpha_<$ but not another satisfies $b - p \geq 0$, since pairs with $b - p < 0$ are in every $E^\alpha_<$. Therefore $\int_{E^\alpha_<}(b-p)\,d\mu_+$ is weakly increasing in $E^\alpha_<$. Also, it is negative when $\alpha = 0$, and is left-continuous. Let $\overline{\alpha}$ be the supremum of values for which $\int_{E^\alpha_<}(b-p)\,d\mu_+ < 0$.

Similarly, $\int_{E^\alpha_>}(p-s)\,d\mu_+$ is weakly decreasing in α, negative at $\alpha = 1$, and right-continuous. Let $\underline{\alpha}$ be the infimum of values for which $\int_{E^\alpha_>}(p-s)\,d\mu_+ < 0$.

We show that $\overline{\alpha} \geq \underline{\alpha}$. Suppose not. Then $E^{\overline{\alpha}}_\leq = (E^{\overline{\alpha}}_< \cup E^{\overline{\alpha}}_=)$ is disjoint from $E^{\underline{\alpha}}_\geq = (E^{\underline{\alpha}}_= \cup E^{\underline{\alpha}}_>)$. We must have $\int_{E^{\overline{\alpha}}_\leq}(b-p)\,d\mu_+ \geq 0$, otherwise the maximality of $\overline{\alpha}$ would be violated. Similarly, $\int_{E^{\underline{\alpha}}_\geq}(p-s)\,d\mu_+ \geq 0$.

Define two new signed measures by

$$\mu'_B(E) = \mu_B(E) - \mu_+(E \cap E^{\overline{\alpha}}_\leq), \qquad \mu'_S(E) = \mu_S(E) - \mu_+(E \cap E^{\underline{\alpha}}_\geq).$$

Note that μ'_B is nonpositive on $E^{\overline{\alpha}}_<$ and nonnegative on $E^{\overline{\alpha}}_>$, hence

$$\int ((b-p) - \overline{\alpha}(b-s))\,d\mu'_B \geq 0.$$

Similarly

$$\int ((p-s) - (1-\underline{\alpha})(b-s))\,d\mu'_S \geq 0.$$

Then we have

$$0 > \int_{\mathbb{R}^2}(b-p)\,d\mu_B - \int_{E^{\overline{\alpha}}_\leq}(b-p)\,d\mu_+ = \int_{\mathbb{R}^2}(b-p)\,d\mu'_B \geq \overline{\alpha}\int_{\mathbb{R}^2}(b-s)\,d\mu'_B,$$

$$0 > \int_{\mathbb{R}^2}(s-p)\,d\mu_S - \int_{E^{\underline{\alpha}}_\geq}(s-p)\,d\mu_+ = \int_{\mathbb{R}^2}(s-p)\,d\mu'_S \geq (1-\underline{\alpha})\int_{\mathbb{R}^2}(s-p)\,d\mu'_S.$$

So $\int_{\mathbb{R}^2}(b-s)\,d\mu'_B < 0$ and $\int_{\mathbb{R}^2}(b-s)\,d\mu'_S < 0$, and therefore

$$\int_{\mathbb{R}^2}(b-s)\,d(\mu'_B + \mu'_S) < 0.$$

However, $\mu'_B + \mu'_S$ is a nonnegative measure since

$$(\mu'_B + \mu'_S)(E) = \mu_+(E) - \mu_+(E \cap E^{\overline{\alpha}}_{\leq}) - \mu_+(E \cap E^{\alpha}_{\geq}) = \mu_+(E \setminus (E^{\overline{\alpha}}_{\leq} \cup E^{\alpha}_{\geq})) \geq 0$$

for any event E. Since $b - s \geq 0$ μ_+-almost everywhere, we have a contradiction.

So indeed we have $\overline{\alpha} \geq \underline{\alpha}$. If $\overline{\alpha} > \underline{\alpha}$, we can take α to be any number in between and β to be arbitrary. Then define

$$\begin{aligned} \widehat{\mu}_B(E) &= \mu_+(E \cap E^{\alpha}_{<}) + \beta\mu_+(E \cap E^{\alpha}_{=}), \\ \widehat{\mu}_S(E) &= \mu_+(E \cap E^{\alpha}_{>}) + (1-\beta)\mu_+(E \cap E^{\alpha}_{=}). \end{aligned}$$

Now

$$E^{\alpha}_{<} \subseteq E^{\alpha}_{<} \cup E^{\alpha}_{=} \subseteq E^{\alpha'}_{<}$$

for any $\alpha' \in (\alpha, \overline{\alpha})$ readily implies

$$\int_{\mathbb{R}^2} b - p \, d\widehat{\mu}_B = \int_{E^{\alpha}_{\leq}} b - p \, d\mu_+ + \beta \int_{E^{\alpha}_{=}} b - p \, d\mu_+ < 0,$$

and by a similar argument

$$\int_{\mathbb{R}^2} p - s \, d\widehat{\mu}_S < 0.$$

Thus, $(\widehat{\mu}_B, \widehat{\mu}_S)$ is a valid pair. Since $\mu_+ \leq \mu$, it is (α, β)-separated, and evidently $\widehat{\mu}_B + \widehat{\mu}_S = \mu_B + \mu_S$, so we are finished in this case.

We are left with the case $\overline{\alpha} = \underline{\alpha}$. In this case, we fix $\alpha = \overline{\alpha} = \underline{\alpha}$ and repeat the argument with β.

Since $b - p, p - s \geq 0$ everywhere on $E^{\alpha}_{=}$, the expression

$$\int_{E^{\alpha}_{\leq}} (b-p) \, d\mu_+ + \beta \int_{E^{\alpha}_{=}} (b-p) \, d\mu_+ \tag{A.1}$$

is weakly increasing in $\beta \in [0,1]$. Let $\overline{\beta}$ be the supremum of such values for which it is < 0. (If it is ≥ 0 already at $\beta = 0$ then take $\overline{\beta} = 0$.) Note that by continuity in β, (A.1) is in fact ≥ 0 at $\overline{\beta}$, except in the corner case where $\overline{\beta} = 1$ and $\alpha = 1$. But we can rule out this corner case where (A.1) is < 0, since in this case we can take $(\alpha, \beta) = (1, 1)$ and the conclusion of the lemma holds — $\int (b-p) \, d\widehat{\mu}_B < 0$ by assumption, $\int (p-s) \, d\widehat{\mu}_S$ must be < 0 because $\widehat{\mu}_S$ only places weight on $E^1_{>}$, where $p - s < 0$ for sure.

Similarly, the expression $\int_{E^{\alpha}_{\geq}} (p-s) \, d\mu_+ + (1-\beta) \int_{E^{\alpha}_{=}} (p-s) \, d\mu$ is decreasing in β; let

$\underline{\beta}$ be the infimum of values for which it is < 0, or $\underline{\beta} = 1$ if no such values exist. The expression is ≥ 0 there except if $\underline{\beta} = 0$ and $\alpha = 0$, and again this corner case can be ruled out.

Now we show that $\overline{\beta} > \underline{\beta}$. Suppose not. Then take any β with $\overline{\beta} \leq \beta \leq \underline{\beta}$. Define

$$\begin{aligned}\mu'_B(E) &= \mu_B(E) - \mu_+(E \cap E^\alpha_<) - \beta\mu_+(E \cap E^\alpha_=),\\ \mu'_S(E) &= \mu_S(E) - \mu_+(E \cap E^\alpha_>) - (1-\beta)\mu_+(E \cap E^\alpha_=).\end{aligned}$$

As before, μ'_B is nonpositive on $E^\alpha_<$ and nonnegative on $E^\alpha_>$, hence

$$\int ((b-p) - \alpha(b-s))\, d\mu'_B \geq 0,$$

and similarly

$$\int ((p-s) - (1-\alpha)(b-s))\, d\mu'_S \geq 0.$$

Now

$$0 > \int_{\mathbb{R}^2} (b-p)\, d\mu_B - \left(\int_{E^\alpha_\leq} (b-p)\, d\mu_+ + \beta \int_{E^\alpha_=} (b-p)\, d\mu_+ \right)$$

(since the first integral is negative by assumption, and the expression in parentheses is just (A.1) at β, which is ≥ 0 because we ruled out the corner case)

$$= \int_{\mathbb{R}^2} (b-p)\, d\mu'_B \geq \alpha \int_{\mathbb{R}^2} (b-s)\, d\mu'_B.$$

Thus, $\int_{\mathbb{R}^2} (b-s)\, d\mu'_B < 0$. By a similar argument, $\int_{\mathbb{R}^2} (b-s)\, d\mu'_S < 0$. Adding, $\int_{\mathbb{R}^2} (b-s)\, d(\mu'_B + \mu'_S) < 0$. But $\mu'_B + \mu'_S$ is identically zero — a contradiction.

Thus, $\overline{\beta} > \underline{\beta}$. So we can choose $\beta \in (\underline{\beta}, \overline{\beta})$. Now let $(\widehat{\mu}_B, \widehat{\mu}_S)$ be defined by (A.1-A.1). It is immediate that $\int_{\mathbb{R}^2} (b-p)\, d\widehat{\mu}_B$, which is just (A.1), is < 0, and similarly $\int_{\mathbb{R}^2} (p-s)\, d\widehat{\mu}_S < 0$. Thus the new pair is valid, and the rest is checked as before. □

Proof of Lemma 3.5: We only prove the formula for Y_B; the Y_S case is analogous.

First suppose $\gamma^*_B = \infty$. Then $\int_{\mathbb{R}^2} w(b,s)\, d\overline{\mu}_B$ is clearly an upper bound for $Y(\alpha,\beta)$. From the definition of γ^*_B, we have $\int_{F^\infty_<} (b-p)\, d\overline{\mu}_B \leq 0$, where $F^\infty_<$ is the event ($w(b,s) > 0$ or $b - p < 0$). If the inequality is strict, we can take $\mu_B = \overline{\mu}_B|_{F^\infty_<}$. Otherwise, since there is a positive probability of $b - p < 0$ under μ (by assumption) and so also under $\overline{\mu}_B|_{F^\infty_<}$ (note that this equals μ for events where $b - p < 0$), then there is also a positive probability of $b - p > 0$ under $\overline{\mu}_B|_{F^\infty_<}$. So we can form μ_B from $\overline{\mu}_B|_{F^\infty_<}$ by removing an

arbitrarily small probability mass on such an event. In either case, we obtain μ_B with $\int_{\mathbb{R}^2}(b-p)\,d\mu_B < 0$ strictly, and $\int_{\mathbb{R}^2} w(b,s)\,d\mu_B$ arbitrarily close to $\int_{\mathbb{R}^2} w(b,s)\,d\overline{\mu}_B$.

Now suppose γ_B^* is finite. Define the measure $\widehat{\mu}_B$ by

$$\widehat{\mu}_B(E) = \overline{\mu}_B(E \cap F_<^{\gamma_B^*}) + \delta_B^* \overline{\mu}(E \cap F_=^{\gamma_B^*}).$$

So the expression given as the value of $Y(\alpha,\beta)$ in the lemma statement is simply $\int w(b,s)\,d\widehat{\mu}_B$. Also, we know that $\int b - p\,d\widehat{\mu}_B = 0$.

We first show that this value is an upper bound on $Y(\alpha,\beta)$. Otherwise, let μ_B be a measure with higher value of $\int w(b,s)\,d\mu_B$, still satisfying $\int(b-p)\,d\mu_B < 0$. Define a signed measure by $\mu'_B = \mu_B - \widehat{\mu}_B$. Then μ'_B is nonpositive on $F_<^{\gamma_B^*}$, and nonnegative on $F_>^{\gamma_B^*}$ (which we define in the obvious way). Therefore,

$$\int_{\mathbb{R}^2} (b-p) - \gamma_B^* w(b,s)\,d\mu'_B \geq 0.$$

This implies

$$\int_{\mathbb{R}^2}(b-p)\,d\mu_B - \int_{R^2}(b-p)\,d\widehat{\mu}_B \geq \gamma_B^*\left(\int_{\mathbb{R}^2} w(b,s)\,d\mu_B - \int_{\mathbb{R}^2} w(b,s)\,d\widehat{\mu}_B\right).$$

But here the left side is negative, while the right side is positive — a contradiction.

So $\int_{\mathbb{R}^2} w(b,s)\,d\widehat{\mu}$ is indeed an upper bound on $Y(\alpha,\beta)$. For the reverse direction, note that, as in the $\gamma_B^* = \infty$ case, the measure $\widehat{\mu}_B$ places some positive probability on the event $b-p<0$ (which is contained in $F_<^{\gamma_B^*}$), and so it must also place positive probability on $b-p>0$. By removing an arbitrarily small amount of probability mass with $b-p>0$, we get a new measure μ_B such that $\int_{\mathbb{R}^2}(b-p)\,d\mu_B < 0$, and $\int_{\mathbb{R}^2} w(b,s)\,d\mu_B$ is arbitrarily close to $\int_{\mathbb{R}^2} w(b,s)\,d\widehat{\mu}_B$. □

References

[1] George A. Akerlof (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (3), 488–500.

[2] George-Marios Angeletos and Jennifer La'O (2013), "Sentiments," *Econometrica* 81 (2), 739–779.

[3] Dirk Bergemann, Benjamin Brooks, and Stephen Morris (2013), "Extremal Information Structures in First Price Auctions," Princeton Economic Theory Center Working Paper #055–2013.

[4] Dirk Bergemann, Benjamin Brooks, and Stephen Morris (2013), "The Limits of Price Discrimination," Princeton Economic Theory Center Working Paper #052–2013.

[5] Dirk Bergemann and Stephen Morris (2013), "The Comparison of Information Structures in Games: Bayes Correlated Equilibrium and Individual Sufficiency," Princeton Economic Theory Center Working Paper #054–2013.

[6] Dirk Bergemann and Stephen Morris (2013), "Robust Predictions in Games with Incomplete Information," *Econometrica* 81 (4), 1251–1308.

[7] Oliver Hart and John Moore (1988), "Incomplete Contracts and Renegotiation," *Econometrica* 56 (4), 755–785.

[8] Nathaniel Hendren (2013), "Private Information and Insurance Rejections," *Econometrica* 81 (5), 1713–1762.

[9] Atsushi Kajii and Stephen Morris (1997), "The Robustness of Equilibria to Incomplete Information," *Econometrica* 65 (6), 1283–1309.

[10] Emir Kamenica and Matthew Gentkow (2011), "Bayesian Persuasion," *American Economic Review* 101 (6), 2590–2615.

[11] Anke S. Kessler (2001), "Revisiting the Lemons Market," *International Economic Review* 42 (1), 25–41.

[12] Anton Kolotilin (2013), "Experimental Design to Persuade," unpublished paper, University of New South Wales.

[13] Jonathan Levin (2001), "Information and the Market for Lemons," *RAND Journal of Economics* 32 (4), 657–666.

[14] Stephen Morris and Hyun Song Shin (2012), "Contagious Adverse Selection," *American Economic Journal: Macroeconomics* 4 (1), 1–21.

[15] Roger B. Myerson and Mark A. Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 29 (2), 265–281.

[16] Luis Rayo and Ilya Segal (2010), "Optimal Information Disclosure," *Journal of Political Economy* 118 (5), 949–987.

[17] Ariel Rubinstein (1989), "The Electronic Mail Game: Strategic Behavior under 'Almost Common Knowledge,'" *American Economic Review* 79 (3), 385–391.