An Online Learning Perspective on Bilateral Trade

MARTINO BERNASCONI
Bocconi University
and
FEDERICO FUSCO
Sapienza University of Rome

Bilateral trade models the mediation between two strategic agents—a buyer and a seller—wishing to exchange a good. While both economists and computer scientists have extensively studied this problem in the Bayesian setting, recent efforts have increasingly focused on mechanisms that operate *without* prior knowledge of the agents' valuations. This article surveys these recent advancements with a focus on *online learning*. We present the literature along three primary axes: gain-from-trade maximization, profit maximization, and related models. Beyond summarizing state-of-the-art results, we aim to build intuition for the rich technical toolkit emerging from this research—techniques we believe will prove valuable well beyond the specific domain of bilateral trade.

1. INTRODUCTION

In the bilateral trade problem, two strategic agents—a seller and a buyer—wish to exchange a good. Both agents hold private valuations for the item and seek to maximize their respective quasi-linear utilities, while the task of designing a mechanism to reach agreement is usually delegated to a third party. This scenario arises naturally in various digital platforms, such as ride-sharing services like Uber or Lyft, where trades between sellers (drivers) and buyers (riders) are managed by a centralized mechanism.

An extensive line of research, building on the seminal work of Myerson and Satterthwaite [1983], has investigated this problem from a Bayesian perspective: the agents' valuations are drawn from known and independent distributions, and the mechanism designer aims to maximize either efficiency or profit while enforcing incentive compatibility and maintaining budget balance (i.e., without subsidizing the agents). The Bayesian approach relies heavily on the assumption that agents draw their valuations from independent distributions whose exact laws are known to the mechanism designer. In this paper, we survey a recent body of work that relaxes this assumption by adopting a machine learning perspective, addressing the following central question:

Under which conditions and how fast is it possible to learn a good mechanism for bilateral trade?

In particular, we focus on the online learning framework introduced by Cesa-Bianchi et al. [2021], where a good mechanism must be learned *on the fly* while minimizing the regret incurred from mistakes made along the way. Beyond summarizing the core results of this literature, this survey also aims to convey the fundamental technical

Authors' addresses: `martino.bernasconi@unibocconi.it`, `fuscof@diag.uniroma1.it`

tools developed in this space, which we believe are of general interest to the broader community.

The remainder of this survey is organized as follows. First, Section 2 formally introduces the bilateral trade problem and its online learning formulation. Next, Sections 3 and 4 present the primary results and techniques for gain-from-trade and profit maximization, respectively. Section 5 then reviews alternative models that have stemmed from the foundational online learning protocol. Throughout the text, we highlight prominent open problems and promising directions for future research.

2. BILATERAL TRADE

We start by introducing the "one-shot" version of bilateral trade, where the agents' valuations are fixed but private. Note that the main focus of this survey is the online learning version of this model, in which a learner is presented with a sequence of T instances of this problem.

In the (static) bilateral trade problem, a seller and a buyer wish to trade a good via an intermediary. The seller holds a single item and is willing to trade it for at least $v_s \in [0,1]$, while the buyer is willing to pay at most $v_b \in [0,1]$ in exchange. The valuations of the two agents are private, so they may misreport them strategically; indeed, they submit bids $(b_s, b_b) \in [0,1]^2$ (not necessarily truthfully) to an intermediary who decides whether the trade happens and under what conditions. More precisely, the intermediary runs a mechanism M, characterized by an allocation region $A \subseteq [0,1]^2$, and pricing rules $p, q : [0,1]^2 \to [0,1]$. The trade happens if and only if the bids (b_s, b_b) belong to the allocation region A, while the payments are made according to p and q. For simplicity, we require $p(b_s, b_b) = q(b_s, b_b) = 0$ whenever $(b_s, b_b) \notin A$, i.e., there is no trade.

The agents' utilities have the standard quasi-linear structure: the buyer's and seller's utilities with valuations v_b and v_s under bids $(b_s, b_b) \in [0,1]^2$ are:

$$U_s\left(b_s, b_b\right) = v_s - \mathbb{1}_{\{(b_s,b_b)\in A\}} \cdot v_s + p(b_s, b_b) \qquad \text{(Seller's utility)}$$
$$U_b\left(b_s, b_b\right) = \mathbb{1}_{\{(b_s,b_b)\in A\}} \cdot v_b - q(b_s, b_b). \qquad \text{(Buyer's utility)}$$

We focus on dominant-strategy incentive compatible (DSIC) and individually rational (IR) mechanisms: each agent maximizes its utility by truthfully reporting its actual private valuation—regardless of the other player's bid—and the utility from participating in the mechanism is at least as high as that from not participating in the mechanism. In formulae,

$$\begin{array}{lll} \text{DSIC:} & U_s\left(v_s, b_b\right) \geq U_s\left(b_s, b_b\right) & \forall\ v_s \in [0,1], (b_s, b_b) \in [0,1]^2 \\ & U_b\left(b_s, v_b\right) \geq U_b\left(b_s, b_b\right) & \forall\ v_b \in [0,1], (b_s, b_b) \in [0,1]^2 \\ \text{IR}: & U_s\left(v_s, b_b\right) \geq v_s,\ U_b\left(b_s, v_b\right) \geq 0 & \forall\ (v_s, v_b) \in [0,1]^2, (b_s, b_b) \in [0,1]^2 \end{array}$$

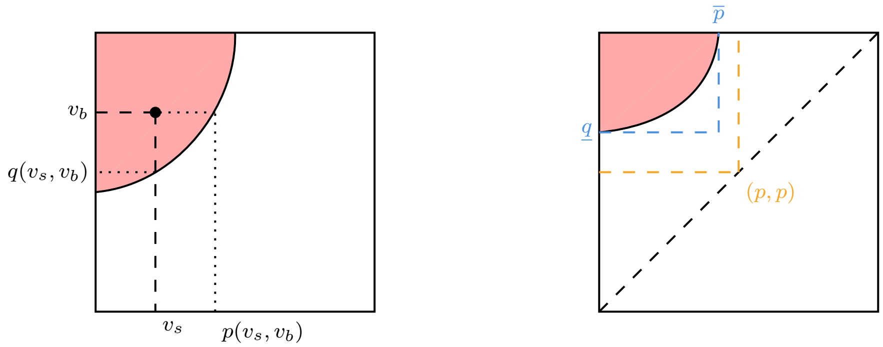


Fig. 1: Visualization of Proposition 2.2 (left) and Proposition 2.5 (right).

2.1 Characterization of DSIC and IR Mechanisms.

Standard mechanism design arguments prescribe that every (deterministic)[1] DSIC and IR mechanisms have a distinctive structure: the payments are uniquely induced by the allocation region, which, in turn, has to respect a monotonicity property. We denote the class of all DSIC and IR mechanisms by $\mathcal{M}$ and restrict to it throughout the rest of the paper, so that we consistently assume that the players' bids are equal to their respective valuations.

Definition 2.1 Monotone Regions & Myerson Payments. *An allocation region $A \subseteq [0,1]^2$ is monotone if for any $x = (x_1, x_2) \in A$ and $y = (y_1, y_2) \in [0,1]^2$ with $x_1 \geq y_1$ and $x_2 \leq y_2$, it holds that $y \in A$. A mechanism is monotone if its allocation region is monotone. Given a monotone allocation region, the associated Myerson payments are defined as follows:*

$$\begin{aligned} p(v_s, v_b) =& \mathbb{1}_{\{(v_s,v_b)\in A\}} \cdot \max\{x \in [0,1] : (x, v_b) \in A\} \qquad \forall\, (v_s, v_b) \in [0,1]^2 \\ q(v_s, v_b) =& \mathbb{1}_{\{(v_s,v_b)\in A\}} \cdot \min\{y \in [0,1] : (v_s, y) \in A\} \qquad \forall\, (v_s, v_b) \in [0,1]^2 \end{aligned}$$

In other words, for an allocation region to be monotone, it should be "closed" in a north-west direction. That is, if a point (v_s, v_b) is in A then any point (v'_s, v_b) with $v'_s \leq v_s$ and any point (v_s, v'_b) with $v'_b \geq v_b$ should also be in A. The payments of a point $(v_s, v_b) \in A$, in turn, correspond to the "east" projection minus the "south" one onto the allocation boundary. We refer to Figure 1 (left) for visualization. Note that for the above definition to be well-posed, we require (without loss of generality) all allocation regions to be topologically closed. By standard mechanism-design arguments [Myerson 1981; Myerson and Satterthwaite 1983], monotone allocation regions and Myerson payments do characterize $\mathcal{M}$ (explicit proof is provided in the Appendices of Di Gregorio et al. [2025]).

Proposition 2.2. *A mechanism for bilateral trade is dominant-strategy incentive compatible and individually rational if and only if its allocation region is monotone and implements Myerson payments.*

[1] Given the *deterministic* nature of the constraints we consider in this survey, we restrict without loss of generality to deterministic mechanisms. Indeed, once the agents' valuations are fixed, any randomized DSIC and IR mechanism can be written as a distribution over deterministic DSIC and IR mechanisms.

2.2 Gain From Trade, Profit, and Budget Balance

Proposition 2.2 describes how a mechanism for bilateral trade should look for the agents to participate truthfully in it. Switching perspectives, we present the most common metrics for assessing a mechanism's quality from the intermediary's perspective: social welfare and profit.

The social welfare induced by a mechanism is equal to the valuation of the agent holding the good *in the end*. More precisely, the social welfare obtained by a generic mechanism $M \in \mathcal{M}$ with allocation region A, when the underlying valuations are (v_s, v_b) is

$$\mathsf{SW}(M, v_s, v_b) = v_b \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}} + v_s \cdot \mathbb{1}_{\{(v_s,v_b)\notin A\}}.$$

An important metric related to social welfare is the gain from trade, which measures the *increase* in social welfare relative to the initial state in which the seller holds the good. In formula, the gain from trade is

$$\mathsf{GFT}(M, v_s, v_b) = (v_b - v_s) \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}}.$$

Note these two metrics are just an additive factor away from each other, as $\mathsf{SW}(M, v_s, v_b) = \mathsf{GFT}(M, v_s, v_b) + v_s$. Social welfare and gain from trade both measure *economic efficiency*, and they are maximized at the same mechanism (while from the "multiplicative" approximation perspective, the latter is more challenging than the former).

The profit extracted by the intermediary from a mechanism $M \in \mathcal{M}$ when the underlying valuations are (v_s, v_b) is simply the difference between the price paid *by* the buyer and the amount paid *to* the seller. In formula:

$$\mathsf{profit}(M, v_s, v_b) = (q(v_s, v_b)) - p(v_s, v_b)) \cdot \mathbb{1}_{\{(v_s,v_b)\in A\}}.$$

So far, it seems there is nothing special about bilateral trade within the realm of mechanism design. However, there is a crucial catch, which was already highlighted in the seminal paper by Vickrey [1961]. Let's say you want to design a mechanism to maximize social welfare, while still enforcing incentive compatibility and individual rationality. Then it makes sense to give the good to the buyer as soon as $v_b \geq v_s$.

Example 2.3. *Consider running the mechanism that allocates in the triangle* $\{(v_s, v_s) \in [0,1]^2 : v_b \geq v_s\}$. *If the agents' valuations are* $(v_s, v_b) = (1/3, 2/3)$, *then the trade happens but induces a* negative profit *for the intermediary, as it pays* $2/3$ *to the seller and receives only* $1/3$ *from the buyer.*

Stated differently, *perfect efficiency may require the mechanism designer to lose money.* To avoid situations like this, where the intermediary *subsidizes* the market, the economic literature has introduced the budget balance constraint, which requires that a social-welfare-oriented mechanism should not incur a negative profit [Myerson and Satterthwaite 1983].

Definition 2.4. *A mechanism* $M \in \mathcal{M}$ *respects the budget balance condition if* $\mathsf{profit}(M, v_s, v_b) \geq 0$, *for all* $(v_s, v_b) \in [0,1]^2$.

In some cases, one may want to require an even stronger condition, so that the profit extracted by the mechanism is exactly 0. This latter condition is called *strong* budget balance (while the one in the above definition is sometimes referred to as *weak* budget balance). Surprisingly, the subfamily of $\mathcal{M}$ that enforces strong budget balance has a very simple characterization. Consider a generic mechanism $M \in \mathcal{M}$, and let $\overline{p}$, respectively $\underline{q}$, be the price paid to the seller, respectively by the buyer, under valuations $(0,1)$; then M enforces budget balance if and only if $\underline{p} \leq \overline{q}$.[1]

A crucial role is played by the family of *fixed-price-mechanisms*. A mechanism $M_{p,q}$ belongs to this family if its allocation region is a rectangle of the form $[0,p] \times [q,1]$, i.e., if it posts price p to the seller and q to the buyer and the trade happens if and only if both agents accept their price. If the two prices coincide, i.e., $p = 1$, then we omit one of the two prices and use M_p to denote it.

Fixed-price mechanisms enjoy the desirable properties of being simple and requiring minimal communication between the intermediary and the agents. Moreover, they somewhat characterize budget-balance mechanisms.[2] More precisely, fixed-price mechanisms are welfare-optimal among all DSIC mechanisms enforcing budget balance and individual rationality.

PROPOSITION 2.5. *Let $M \in \mathcal{M}$ be any budget-balanced mechanism, and denote with $\overline{p}$ and $\underline{q}$ the prices corresponding to valuation $(0,1)$. Then the following inequalities hold for any $(v_s, v_b) \in [0,1]^2$ and any $p \in [\overline{p}, \underline{q}]$:*

$$\mathsf{SW}(M, v_s, v_b) \leq \mathsf{SW}(M_{\overline{p},\underline{q}}, v_s, v_b) \leq \mathsf{SW}(M_p, v_s, v_b).$$

There is a simple way of picturing the above proposition, given that none of the mechanisms forces a trade with negative gain from trade (i.e., when $v_s > v_b$): the allocation region of the initial mechanism A is contained into the $[0,\overline{p}] \times [\underline{q},1]$ square which, in turns, is contained in the $[0,p] \times [p,1]$ rectangle, for any $p \in [\overline{p}, \underline{q}]$. We refer to Figure 1 (right) for a "proof-by-picture". We conclude by noting that, even though fixed-price mechanisms are welfare-optimal, they are not generally profit-optimal among budget-balanced mechanisms. To see this, consider the following example, where the private valuations are not fixed but drawn from some distribution.

EXAMPLE 2.6. *Consider the uniform distribution over valuations in the $(0, 1/2)$–$(\delta, 1)$ segment, where δ is an arbitrarily small parameter. The mechanism allocating in the $(0, 1/2)$–$(\delta, 1)$–$(0,0)$ triangle would extract an expected profit of $\approx 3/4$, while any fixed price mechanism would not extract more than $\approx 1/2$.*

2.3 A Quick Digression: The Bayesian Perspective

Although this is a survey on bilateral trade and online learning, we devote some quick words to describe the Bayesian perspective and the corresponding results. In that line of work, the agents' private valuations are drawn *once* from some independent and *known* distributions. First, we recall that the seminal paper by Myerson and Satterthwaite [1983] shows that incentive compatibility, individual rationality, and budget balance are incompatible with full efficiency. Actually, this

[1] One implication is easy, since a budget balanced M should not lose money for valuations $(0,1)$. The other implication follows by monotonicity. The price $\underline{p}$ upper bounds the payments to the seller under *any* valuation in the allocation region, and similarly $\overline{q}$ lower bounds the payments made by the buyer in the same situation.

[2] Already Hagerty and Rogerson [1987] showed that fixed-price mechanisms that post a single price to both agents are the only possible DSIC, IR, and strong budget-balanced mechanisms.

result holds for the weaker notions of (Bayesian) incentive compatibility, (interim) individual rationality, and budget balance which only need to hold *in expectation*.

Given this impossibility result, the most natural question that arises is whether a constant *multiplicative* factor of perfect efficiency is attainable while enforcing the other constraints. For social welfare, Blumrosen and Dobzinski [2021] provide a $1-1/e$ approximation, later improved to $1 - 1/e + 10^{-4}$ in Kang et al. [2022]. The current state of the art is that the optimal approximation factor achievable is contained in the $[0.72, 0.7381]$ interval [Cai and Wu 2023; Liu et al. 2023]. Interestingly, it turns out that fixed-price mechanisms are insufficient to extract a constant factor of the optimal gain from trade [Blumrosen and Mizrahi 2016]. Instead, the random-offerer mechanism, which randomly delegates the pricing power to the seller or the buyer, can provide such a guarantee [Deng et al. 2025] (note, the random-offerer mechanism is no longer DSIC, but only ensures truthfulness *in expectation*). The state of the art places the optimal gain-from-trade approximation achievable in the $[1/3.15, 2/e]$ interval [Fei 2022; Blumrosen and Mizrahi 2016]. We also mention a recent paper [Hajiaghayi et al. 2025] studying the interplay between gain from trade and profit maximization, investigating how much gain from trade a profit-maximizing intermediary can ensure.

We finally note that the online learning approach is not the only approach adopted in the literature to overcome the Bayesian assumption (i.e., *independent* agents' distributions that are *perfectly known* to the mechanism designer). For instance, a line of work has investigated the number of samples (either one or more) needed to get a meaningful multiplicative approximation to the optimal mechanisms [Kang et al. 2022; Hajiaghayi et al. 2025; Dütting et al. 2026], while Dobzinski et al. [2025] investigate the one-shot problem when the agents' valuations may exhibit correlation.

2.4 The Learning Protocol

We are ready to introduce the main character of the survey: the learning protocol for bilateral trade, originally introduced in the conference version of Cesa-Bianchi et al. [2024a]. At each time step t, a new pair of agents arrives, with private valuations (v_s^t, v_b^t), while the learner/intermediary proposes a mechanism $M^t \in \mathcal{M}$. The trade then happens according to the valuations and the mechanism, with the learner observing some feedback z_t. We refer to the pseudocode for further details. For simplicity, we use $\mathsf{GFT}_t(M)$, respectively $\mathsf{profit}_t(M)$, to denote the gain from

The Learning Protocol

for time $t = 1, 2, \ldots$ **do**
  a new pair of agents arrives with private valuations $(v_s^t, v_b^t) \in [0, 1]^2$
  the learner declares a mechanism $M^t \in \mathcal{M}$
  the trade takes place according to M^t and (v_s^t, v_b^t)
  the learner observes some feedback z_t
  the learner is awarded either $\mathsf{GFT}(M^t, v_s^t, v_b^t)$ or $\mathsf{profit}(M^t, v_s^t, v_b^t)$

trade, respectively profit, of the generic mechanism M on the valuations (v_s^t, v_b^t)

at time t. The learning protocol offers many degrees of freedom beyond the profit vs. gain-from-trade dichotomy. We list here some of them:

- **Feedback Models.** There are many meaningful feedback models. For instance, for a general mechanism, it makes sense to directly observe both valuations, as they are needed to compute Myerson payments. If we restrict ourselves to fixed-price mechanisms, however, then a simple yes/no answer to the proposed prices may be enough.
- **Data-Generation Models.** It is possible to study many adversaries generating the sequence of valuations. Beyond the standard stochastic i.i.d. and the adversarial ones, an interesting model is provided by the σ-smooth adversary.
- **Budget Balance Constraints.** In gain-from-trade maximization scenarios, it makes sense to investigate the impact of budget balance constraints on the learnability of the problem. Beyond strong and weak *per-round* budget balance, the literature has also investigated more global versions, such as constraining the overall (i.e., cumulative) negative profit.

As is standard in online learning, a learning mechanism is measured by its *regret* relative to a fixed benchmark. In formula, the regret of a gain-from-trade maximizing algorithm $\mathcal{A}$ against an adversary $\mathcal{S}$ is

$$R_T(\mathcal{A},\mathcal{S}) = \sup_{M\in\mathcal{M}_{\mathrm{BB}}} \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{GFT}_t(M^t) - \mathsf{GFT}_t(M)\right],$$

where the algorithm chooses budget-balanced mechanisms $M^t \in \mathcal{M}$, while the adversary generates the valuation pairs (v_s^t, v_b^t) (and thus $\mathsf{GFT}_t(\cdot)$). Here we use $\mathcal{M}_{\mathrm{BB}}$ to stress that the benchmark is in $\mathcal{M}$ and is also restricted to respect budget balance. In practice, Proposition 2.5 tells us that the benchmark is the best fixed price mechanism for $\mathcal{S}$. The regret of algorithm $\mathcal{A}$ against a certain family of adversaries (e.g., stochastic i.i.d) is denoted as the worst case regret within that family: $R_T(\mathcal{A}) = \sup_{\mathcal{S}} R_T(\mathcal{A},\mathcal{S})$. Note that gain from trade and social welfare are identical up to an additive v_s; therefore, studying either of the two is perfectly equivalent in the regret minimization world.

For profit-maximizing algorithms, the definition is analogous, with the sole difference that there is no need to explicitly restrict the algorithm and the benchmark to budget-balanced mechanisms:

$$R_T(\mathcal{A},\mathcal{S}) = \sup_{M\in\mathcal{M}} \mathbb{E}\left[\sum_{t=1}^{T} \mathsf{profit}_t(M^t) - \mathsf{profit}_t(M)\right].$$

3. GAIN FROM TRADE

In this section, we present the line of research relative to gain-from-trade-maximizing learning algorithms. Motivated by Proposition 2.5, these works focus on fixed-price mechanisms, which is particularly interesting from the online learning perspective. Indeed, fixed-price mechanisms allow for peculiar feedback models: when the learner/intermediary posts two prices, one for the seller and one for the buyer, it is natural to only observe whether the agents accept or not, without eliciting their actual valuations. The literature mainly studies two feedback models: the one-bit model, where the learner only observes whether the trade happens or not, and the two-bit one, where the learner separately observes the answers of both agents.[1]

In the presentation, we put more emphasis on the most challenging (and arguably natural) setting: one-bit feedback and non-stationary valuations under different budget-balance conditions. Beyond these two *partial* feedback models, the literature has also investigated the paradigmatic full feedback setting as a reference.

We introduce some simplified notation for the fixed-price setting: At each time t, the learner proposes a price p_t to the seller (with valuation v_s^t) and a price q_t to the buyer (with valuation v_b^t). Then the gain from trade simplifies to

$$\mathsf{GFT}_t(p_t, q_t) = (v_b^t - v_s^t)\mathbb{1}_{\{v_s^t \le p_t \wedge q_t \le v_b^t\}},$$

Similarly, the profit becomes $\mathsf{profit}_t(p_t, q_t) = (q_t - p_t)\mathbb{1}_{\{v_s^t \le p_t \wedge q_t \le v_b^t\}}$. Given Proposition 2.5, it seems redundant to study posted price mechanisms with two different prices. However, one surprising result of this line of work is that posting two prices actually helps in learning settings.

3.1 Technical Challenges of GFT Maximization

We start by presenting two features that differentiate GFT maximization from other online learning settings.

The first one is a "lack of observability" under partial feedback (i.e., one- or two-bit feedback). While the performance of a learner is evaluated in terms of the total gain from trade accumulated, such an objective is never observed *directly* by the learner, as the valuations v_s, v_b remain hidden even if the trade is accepted by both parties. For instance, if the learner posts price $p = 1/2$ to both agents and observes that the trade is accepted, it cannot discriminate between the case in which $v_s = 0, v_b = 1$ (and thus the GFT is 1) and the case in which $v_s = v_b = 1/2$ (and thus the GFT is 0).

Second, the action space $[0, 1]^2$ is continuous, and the reward is highly irregular, i.e., it is neither continuous, one-sided Lipschitz nor concave. This might lead to the so-called "needle in a haystack" problem. For instance, consider an adversary selecting an hidden "needle" x arbitrarily close to $1/2$ and sampling valuations uniformly and independently at random from $\{(0, x), (x, 1)\}$. Then, posting any price $p \neq x$ leads to a GFT of $\approx 1/4$. However, in this instance, the optimal price is exactly the needle x, which leads to a GFT of $\approx 1/2$.

3.2 Strong Budget Balance

The strong budget balance condition requires that the learner posts the same price to both agents, at each time step, i.e., $p_t = q_t$ for all $t = 1, \dots, T$. This means that the platform does not extract any profit from the market, nor does it subsidize it, which is reasonable in many contexts that require a platform that only guarantees facilitating the trade between participants.

With one bit of feedback at each time step, posting one single price is not enough

[1] In principle, we could also learn a fixed-price mechanism by employing more general mechanisms and committing only to a fixed-price one once valuations have been learned; however, this would require a stronger feedback model, which contrasts with the advantage of using fixed-price mechanisms, and, to the best of our knowledge, this hybrid model has not been studied.

	Stationary i.i.d. Settings				Non-Stationary Settings	
	i.i.d.	+ bd	+ ind.	+ bd + ind.	σ-smooth	Adversarial
Full	$\tilde{\Theta}(\sqrt{T})$					$\Omega(T)$
2-bit	$\Omega(T)$			$\tilde{\Theta}(T^{2/3})$	$\Omega(T)$	
1-bit	$\Omega(T)$					

Table I: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **strong budget balance**. The rows correspond to feedback models (full feedback, two- and one-bit feedback), while the columns correspond to data generation models (in particular, bd denotes bounded density, while ind. denotes that the seller and buyer distributions are independent).

for the learner to achieve sublinear regret under one-bit feedback *even against the stochastic i.i.d. adversary*, see Theorem 5 of Cesa-Bianchi et al. [2024a]. The lower bound construction is simple and is based on the "lack of observability" property: it is possible to construct two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ over valuations which induce the same feedback distribution, i.e., for any $p \in [0,1]$, the feedback $z_t = \mathbb{1}_{\{v_s \le p\}} \cdot \mathbb{1}_{\{v_b \ge p\}}$ follows the same law under both $\mathcal{D}_1$ and $\mathcal{D}_2$. At the same time, the optimal prices under the two distributions are far apart, so that any algorithm suffers linear regret because it is forced to take the same actions in both instances, and thus it will incur linear regret in at least one of the two.

We mention some results beyond one-bit feedback due to Cesa-Bianchi et al. [2024a]. In the adversarial setting *even full feedback* is not enough to achieve sublinear regret, while in the stochastic i.i.d. setting no-regret is attainable. In particular, under two-bit feedback, sublinear regret is achievable if and only if the agents' distributions are independent and admit bounded density. Similarly, if the learner has access to full feedback, then no-regret is achievable in the stochastic i.i.d. setting without assuming anything about the distributions.

The bounded density assumption can be generalized in a non-stationary setting using the notion of σ-smooth adversary [Haghtalab et al. 2020], which is a popular beyond-worst-case input assumption [Roughgarden 2020]. In particular, a σ-smooth adversary for bilateral trade chooses at each time step a distribution $\mathcal{D}_t$ over $[0,1]^2$ which admits a density bounded by σ^{-1}.[1]

In a follow-up work, Cesa-Bianchi et al. [2024b] show that no-regret is attainable under this assumption in the full feedback setting (thus yielding a positive result in a non-stationary setting), while partial feedback still remains useless. The exact minimax rates of the known results for strong budget balance learning algorithms are summarized in Section 3.2. For simplicity, here and in the following, we only report the time-horizon dependency in the regret rates, without specifying the impact of the smoothness term $1/\sigma$. Indeed, while the algorithms exhibit regret rates that are polynomial in $1/\sigma$, the known lower-bound rates are independent of this term.

[1] Or, alternatively, if μ denotes the Lebesque measure and $\mathcal{B}([0,1]^2)$ the Borel sigma-algebra of the unit square, $\mathcal{D}_t$ respects that $\mathbb{P}_{v\sim\mathcal{D}}(v \in A) \le \frac{\mu(A)}{\sigma}$, for all $A \in \mathcal{B}([0,1]^2)$.

3.3 Weak Budget Balance

Proposition 2.5 states that, in the (one-shot) Bayesian setting, there is no point in posting two different prices. Surprisingly, from the online learning perspective, this is not the case.

For fixed-price mechanisms, the (weak) budget balance condition requires that the seller price p_t and the buyer price q_t respect $p_t \leq q_t$, thus guaranteeing that $\mathsf{profit}_t \geq 0$ for all t. Although suboptimal from an optimization point of view (posting any price $p \in [p_t, q_t]$ yields at least the same gain-from-trade), it turns out that posting two prices allow to estimate the gain-from-trade with one-bit feedback. More formally, Azar et al. [2024] show that for every target price $p \in [0, 1]$, there exists a randomized procedure that produces two random variables $\hat{p}, \hat{q} \in [0, 1]$ such that the probability of observing the trade is exactly $\mathsf{GFT}(p, v_s, v_b)$ for the underlying hidden valuations (v_s, v_b). In formulas,

$$\mathbb{P}(v_s \leq \hat{p}, \hat{q} \leq v_b) = \mathsf{GFT}(p, v_s, v_b).$$

Algorithm 1 GFT estimator with weakly budget balanced prices

Require: $p \in [0, 1]$

1. With probability p: $\hat{p} = U$, $\hat{q} = p$, where $U \sim \text{Uniform}[0, p]$
2. With probability $1 - p$: $\hat{p} = p$, $\hat{q} = V$, where $V \sim \text{Uniform}[p, 1]$

return $(\hat{p}, \hat{q})$

The simple estimator is shown in Algorithm 1. Having an unbiased estimator of the reward solves the first challenge of GFT maximization with partial feedback, i.e., the lack of observability of the reward. However, this alone does not solve the "needle in a haystack" problem, so that no-regret is still unattainable in the adversarial setting or the stochastic i.i.d. one with either one or two bit feedback (see Cesa-Bianchi et al. [2024b] for the adversarial setting, while the independent lower bound from Cesa-Bianchi et al. [2024a] still holds for weak budget balance). However, it provides a first step towards achieving positive results in a non-stationary environment and partial feedback.

Notably, the estimator only needs one bit of feedback, and has been used in Cesa-Bianchi et al. [2024b] to derive a positive result for the σ-smooth adversary, that we briefly present here. The main property given by σ-smoothness is that it turns the GFT into a Lipschitz function. More formally, Lemma 1 of Cesa-Bianchi et al. [2024b] ensures that $p \mapsto \mathbb{E}[\mathsf{GFT}_t(p)]$ is $1/\sigma$-Lipschitz. This assumption addresses the "needle in a haystack" problem: playing only the discretized prices $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ leads to a discretization error (the difference between the maximum cumulative GFT on $[0, 1]$ and the maximum on the discretized prices) of at most $\sigma^{-1}\varepsilon T$. Then, one can combine the estimator in Algorithm 1 with the block-decomposition algorithm of Awerbuch and Mansour [2003]. The resulting algorithm divides the time horizon T into N blocks, and in each of the N blocks, it assigns randomly $O(1/\varepsilon)$ steps (one for each discrete price) to perform the estimation procedure, and at the end of each block, it updates some full-feedback regret minimization algorithm (e.g., Hedge [Cesa-Bianchi and Lugosi 2006]) and probes from it the next price to play in

	Stationary i.i.d. Settings		Non-Stationary Settings	
	i.i.d.	+ bd	σ-smooth	Adversarial
Partial Feedback	$\Omega(T)$	$\tilde{\Theta}(T^{3/4})$	$\tilde{\Theta}(T^{3/4})$	$\Omega(T)$

Table II: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **budget balance**. Partial feedback refers indifferently to either one-bit or two-bit feedback, while the columns correspond to data generation models (in particular, bd denotes bounded density).

the next block (modulo the estimation steps). Informally, by fixing $\varepsilon \approx T^{-1/4}$ and $N \approx \sqrt{T}$, one can derive the following regret bound:

$$R_T \lesssim \underbrace{\frac{T}{N}}_{\text{scale}} R_N^{\textsc{Hedge}} + \underbrace{\frac{N}{\varepsilon}}_{\substack{\text{estimation}\\\text{rounds}}} + \underbrace{\frac{1}{\sigma}\varepsilon T}_{\substack{\text{discretization}\\\text{error}}} \lesssim O\left(\frac{1}{\sigma}T^{\frac{3}{4}}\right),$$

where we use that the regret of Hedge over $O(1/\varepsilon)$ experts and N time steps is $R_N^{\textsc{Hedge}} \in \tilde{O}(\sqrt{\log(1/\varepsilon)N})$, which gets multiplied for the scale of the rewards at each time block (T/N). It turns out that this rate is essentially tight, even with two-bit feedback. This is due to a lower bound of Cesa-Bianchi et al. [2024b] which holds for constant σ. The known results for weak budget balance algorithms are summarized in Table II; note that in full feedback there is no point in posting different prices to seller and buyer (so that the minimax rates are the same as in Section 3.2).

3.4 Global Budget Balance

The σ-smooth assumption is useful because it solves the "needle in a haystack" problem. Bernasconi et al. [2024b] use another approach to solve this problem, without weakening the adversary, but by relaxing the budget balance constraint. Inspired by the bandit with knapsack (BwK) framework with non-monotone resource consumption [Kumar and Kleinberg 2022; Slivkins et al. 2024; Bernasconi et al. 2024a], they define a *long-term* constraint on the budget, i.e., they require a *global* budget balance condition:

$$B_T \geq 0, \quad \text{where} \quad B_0 = 0, B_{t+1} = B_t + \mathsf{profit}_t(p_t, q_t).$$

Let's see how allowing for a controlled amount of negative profit helps finding the needle in the haystack. For instance, recall the hard distribution which outputs uniformly at random either $(v_s, v_b) = (0, 1/2)$ or $(1/2, 1)$. By only playing budget balanced prices (p, q) with $p \leq q$, only $p = q = 1/2$ has GFT of 1, while any other price has GFT of $1/2$. However, if we allow p to be slightly greater than q (let's say $p = q + \varepsilon$) then playing $(q + \varepsilon, q)$ for any $q \in [1/2 - \varepsilon, 1/2]$ induces a GFT of 1, by only paying ε negative profit. This simple observation leads to a global budget-balanced algorithm with one-bit feedback, adversarial valuations, and $\Theta(T^{3/4})$ regret. The upper bound comes from Bernasconi et al. [2024b] while the tight lower bound is due to two independent and concurrent works: Lunghi et al. [2026a] and [Chen et al. 2025]. The idea of the algorithm is to divide the time horizon into two phases: the first phase allows accumulating a profit budget β, and the second phase allows spending that budget by playing slightly non-budget-balanced prices.

The second phase almost exactly models the weak budget balanced constraint, i.e.,

	Stationary i.i.d. Settings		
	i.i.d.	+ ind.	Adversarial
Full Feedback	$\tilde{\Theta}(\sqrt{T})$		
Partial Feedback	$\tilde{\Theta}(T^{3/4})$	$\tilde{\Theta}(T^{2/3})$	$\tilde{\Theta}(T^{3/4})$

Table III: Minimax rates achievable for **gain-from-trade** maximization in bilateral trade with **global budget balance**. Partial feedback refers indifferently to either one-bit or two-bit feedback, while the columns correspond to data generation models (in particular, ind. denotes that the seller and buyer distributions are independent).

to use an estimator for non-budget-balanced prices $H_K = \{(i/K, (i-1)/K)\}_{i\in[K]}$ and the block decomposition algorithm. The only difference is that the estimator for the prices in H_K is slightly biased (by an additive $O(1/K)$ term). The key technical challenge is to show that, while maximizing the profit in the first phase, the algorithm does not lose too much GFT. Concretely, Bernasconi et al. [2024b, Lemma 4.1] provides a multiplicative discrete grid F_K (of size $\tilde{O}(K)$) of budget-balanced prices and show that

$$\max_{p\in[0,1]} \sum_{t\in[T]} \mathsf{GFT}_t(p) \le O(\log T) \max_{(p,q)\in F_K} \sum_{t\in[T]} \mathsf{profit}_t(p,q) + O\left(\frac{T}{K}\right).$$

This result is crucial in relating the global budget constraint with the (gain-from-trade) regret, ensuring that the profit-maximizing phase does not induce "too much regret". More precisely, this shows that the optimum price on the grid F_K in terms of profit approximates the optimum price in terms of GFT(up to log factors). The first phase is conducted until a budget of β is accrued (call the stopping time τ). In this phase, the algorithm instantiates any regret minimizer over F_K. This let us conclude that the maximum GFT until τ is small, i.e.

$$\max_{p\in[0,1]} \sum_{t\in[\tau]} \mathsf{GFT}_t(p) \le O(\log T)(\beta + \sqrt{TK}) + O\left(\frac{T}{K}\right).$$

Choosing $\beta \in \Theta(T^{3/4})$, $K \in \Theta(T^{1/4})$ and $N \in \Theta(T^{1/2})$, we can show that $B_T \ge 0$ (since the algorithm loses at most $T/K \le \beta$ profit in the second phase), and thus global budget balance is satisfied. In the first phase, with our choice of parameters, $\max_{p\in[0,1]} \sum_{t\in[T]} \mathsf{GFT}_t(p) \in \tilde{O}(T^{3/4})$ and in the second phase we can mimic the calculations done in the weak budget balance section which also give a regret of $\tilde{O}(T^{3/4})$, for a total regret is $O(T^{3/4})$.

Beyond the matching adversarial lower bound, Chen et al. [2025] provides tight results for other data generation models. All known minimax rates are summarized in Table III.

3.5 Other Works on GFT Maximization

Beyond the models presented so far, other works provide a broader view on gain-from-trade maximization in bilateral trade, across several dimensions: budget violations, stronger benchmarks, and intermediate feedback models.

3.5.1 *Other Budget Balance Notions.* The two-phase algorithm of Bernasconi et al. [2024b] first accumulates the profit budget and then spends it by posting mildly subsidizing prices. A natural question is what happens when the learner is allowed to violate the global budget constraint. Lunghi et al. [2026a] study the impact of budget violation allowed and the achievable regret. In particular, they show that for a violation of the order of $O(T^{\beta})$ (i.e., if the learner is allowed to lose up to T^{β} profit throughout the time horizon), with $\beta \in [3/4, 6/7]$, a tight regret of order $\tilde{O}(T^{1-\beta/3})$ can be obtained. On the other hand, for $\beta \leq 3/4$, violations are not useful, since the regret of $\tilde{O}(T^{3/4})$ cannot be improved upon, while also violating more than $\beta \geq 6/7$ does not lead to an improvement since $O(T^{5/7})$ is tight in that regime.

3.5.2 *Other Benchmarks.* The global budget balance constraint motivates stronger benchmarks than the best fixed price. If the agent is allowed to balance profit and losses over the whole horizon, it is natural to compare against the best *distribution* over fixed prices whose expected profit is non-negative. This benchmark can be strictly stronger than the best deterministic price: randomizing across prices may use profitable trades to finance subsidized trades and thereby increase total gain from trade. However, Bernasconi et al. [2024b] shows that they cannot be further than a multiplicative factor of 2, namely that, for any sequence of valuations:

$$\sum_{t \in [T]} \mathbb{E}_{(p,q)\sim\gamma^*}[\mathsf{GFT}_t(p,q)] \leq 2 \sum_{t \in [T]} \mathsf{GFT}_t(p^*),$$

where γ^* is the best feasible distribution in expectation and p^* is the best fixed price. This benchmark is too strong in the adversarial model, indeed, Bernasconi et al. [2024b] show a linear $(1+\Omega(1))$-regret lower bound.[1] The bounded constant multiplicative ratio mentioned above shows that any no-regret algorithm for the single fixed price is a 2-regret algorithm against the best feasible distribution, so there is a gap between the upper bound and lower bound for this particular benchmark and adversarial valuations. This impossibility motivates studying stochastic models.

Lunghi et al. [2026a] shows that stochastic valuations are not enough, and that there is a linear regret lower bound. However, they also show that for a valuation distribution with bounded density, the minimax regret rate is $\tilde{\Theta}(T^{3/4})$. In a subsequent work, Lunghi et al. [2026b] construct an algorithm that scales with the amount of perturbation C, which interpolates between the stochastic (with bounded density), which is the case $C = 0$, to the adversarial one, which is the case $C = T$, and obtains a regret of $\tilde{O}(T^{3/4}) + O(C \log T)$, against the distributional benchmark. Moreover, the algorithm does not need to know the perturbation parameter C, which is defined as the sum over $t \in [T]$ of total variation distance between the distribution choose at time t by the adversary and the unperturbed distribution.

Finally, Azar et al. [2024] studies 2-regret (with respect to the best fixed-price) under different feedback schemes and under the strong and weak budget-balanced conditions. In particular, they prove that sublinear 2-regret is not achievable with one- or two-bit feedback under the strong budget-balanced condition. On the other

[1] For $\alpha > 1$, we define α-regret as the normal regret in which the baseline is divided by α, in the spirit of competitive ratio.

hand, under the weak budget-balanced condition, sublinear 2-regret is achievable even with one-bit feedback.

3.5.2.1 *Other feedback models.* There are other intermediate feedback models for GFT maximization. Bacchiocchi et al. [2025] studies an *asynchronous* mechanism that queries the seller only if the buyer has already accepted the offer, which is a feedback in between the one-bit and two-bit one, as it receives a clean bit from the buyer and a censored one from the seller. Under this feedback model and a strong budget balance, the only known algorithm that achieved this was that of Cesa-Bianchi et al. [2024a], which achieved regret of $\Omega(T^{2/3})$ under the bounded-density assumption. Bacchiocchi et al. [2025] matches this under this more restricted feedback.

Finally, Chen et al. [2025], Jin [2026] and Lunghi et al. [2025] study upper and lower bounds for an intermediate feedback (called semi-feedback in Chen et al. [2025] and Jin [2026], and asymmetric-feedback in Lunghi et al. [2025]) in which one of the two valuations is observed directly, for instance, the buyer's valuation v_b, while for the seller's valuation v_s, we either observe the outcome of the seller's decision or the overall success of the trade.

In the global budget balance regime, they prove a tight $T^{2/3}$ regret bound, improving from the $T^{3/4}$ regret obtainable under just one-bit feedback.

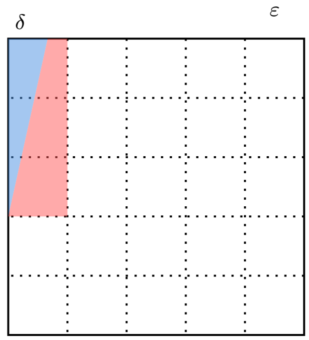
4. PROFIT

In this section, we present the results for profit maximization. Although the profit function resembles GFT, we do not have an equivalent characterization as in Proposition 2.5. This means that we cannot restrict our attention only to fixed-price mechanisms (see Example 2.6) and, to compute Myerson payments, we need the agents to reveal their valuations. In online learning jargon, this means that profit-maximizing algorithms receive *full feedback*. Compared to gain-from-trade maximization, we then have an incredibly larger action space (all monotone allocation regions vs. all fixed prices) but a more informative feedback (full vs. one- or two-bit). These two structural differences call for new technical tools.

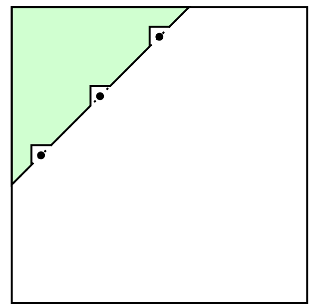
When faced with a large action space, the natural approach in online learning is to discretize it (e.g., in Lipschitz bandits [Kleinberg et al. 2019] or pricing [Kleinberg and Leighton 2003]). This method splits the problem into two parts: first, finding *offline* a finite subset of actions that well approximates the optimal one *uniformly over all possible inputs*, and then running *online* a discrete learning algorithm on the discretized action space. The most natural discretization of the mechanism space $\mathcal{M}$ is to cast a uniform ε-grid on the $[0,1]^2$ square and consider all the $\binom{2/\varepsilon}{1/\varepsilon}$ mechanisms $\mathcal{M}_\varepsilon^\perp$ with monotone allocation regions that can be described as the union of the cells of the grid[1]. Unfortunately, as we show in the following example, this rich class of mechanisms does not provide a uniform approximation for our problem. This is due to the non-continuity (and thus non-Lipschitzness) of our objective function.

Example 4.1 (Fixed Discretizations Fail). *Consider the uniform distribution $\mathcal{D}$ over the segment $(0, 1/2)$-$(\delta, 1)$, where δ is an arbitrarily small parameter*

[1]The family $\mathcal{M}_\varepsilon^\perp$ is a uniform ε-coverage of the mechanism space, for a suitable definition of distance between mechanisms.



(a) Visualization of Example 4.1.



(b) Hard Instance of Example 4.2.

Fig. 2: Supporting figures for the hard examples.

(as in Example 2.6). The corresponding optimal mechanism M allocates in the $(0, 1/2)$-$(\delta, 1)$-$(0, 1)$ triangle, for an expected profit of approximately $3/4$. Consider now any fixed discretization parameter $\varepsilon > \delta$, the support of $\mathcal{D}$ is contained in the first column of the grid, so the profit extracted by any mechanism $M \in \mathcal{M}_\varepsilon^\perp$ only depends (up to $O(\delta)$ payment to the seller) on the horizontal segment that crosses such column, for a maximum profit of $1/2$ (see also Figure 2a for a visualization).

This simple construction is at the core of the adversarial lower-bound construction in Di Gregorio et al. [2025], showing that no learning algorithm can achieve sublinear regret when the task is maximizing profit in the adversarial setting. In the same paper, the authors study the problem in the stochastic i.i.d. setting and show that learning is indeed possible there, with a regret of $\tilde{O}(\sqrt{T})$, which is tight up to poly-logarithmic terms. Finally, Di Gregorio et al. [2026] complete the picture of natural data-generation models, proving that nearly tight $\tilde{O}(\sqrt{T})$ is achievable also against a non-stationary adversary, at the cost of assuming σ-smoothness. The rest of the section is devoted to presenting briefly the challenges and the techniques used to achieve the $\tilde{O}(\sqrt{T})$ regret in the stochastic i.i.d. setting. We refer to Table IV for a summary of these results.

4.1 Profit Maximization against the i.i.d. Stochastic Adversary

The well-known online-to-offline reduction (e.g., Cesa-Bianchi et al. [2004]) shows that finding the best mechanism online in the stochastic i.i.d. setting *under full feedback* is equivalent to studying the standard offline PAC learning framework. In particular, deriving a $\sqrt{T}$ regret bound is equivalent to showing that, given access to approximately $1/\varepsilon^2$ i.i.d. samples from the underlying distribution $\mathcal{D}$, one is capable of finding an $M \in \mathcal{M}$ such that

$$\sup_{\mathcal{M}^\star \in \mathcal{M}} \mathbb{E}\left[\mathsf{profit}(\mathcal{M}^\star, v)\right] \lesssim \mathbb{E}\left[\mathsf{profit}(M, v)\right] + O(\varepsilon), \qquad (1)$$

where v denotes a fresh valuation drawn from $\mathcal{D}$. It turns out that even with this simplification, the task poses significant challenges: the family of "profit functions" $\{\mathsf{profit}(M, \cdot)\}_{M \in \mathcal{M}}$ is statistically extremely complicated! In particular, it exhibits unbounded fat-shattering dimension for constant values of the margins (and thus it also has unbounded pseudo-dimension) [Di Gregorio et al. 2026]. Instead of proving

Stochastic i.i.d. Setting	Smooth Adversary	Adversarial Setting
$\tilde{\Theta}(\sqrt{T})$	$\tilde{\Theta}(\sqrt{T})$	$\Omega(T)$

Table IV: Minimax regret rates achievable for profit-maximizing algorithms in bilateral trade.

these impossibility results, we propose here a convincing and direct exhibit of the complexity of learning this class of mechanisms: it does not enjoy the uniform convergence property. Stated differently, regardless of the number of samples a learner has access to, there may be some mechanisms whose empirical profit is far from the expected one.

Example 4.2 (No Uniform Convergence). *Let $\mathcal{D}$ be the uniform distribution on the $(0, 3/4)$-$(1/4, 1)$ segment, and consider any realization of n i.i.d. samples S from $\mathcal{D}$. Denote with $\delta > 0$ a small parameter we set later, and let M be the mechanism allocating in the $(0, 3/4)$, $(1/4, 1)$, $(0, 1)$ triangle, minus the δ-radius ℓ_∞ balls $B_\delta(v)$ for v in S. By construction, the empirical profit of M on the samples is zero, while the expected profit on $\mathcal{D}$ is at least $1/4$. In fact, S is finite, so it is possible to take δ small enough so that at least half of the support of $\mathcal{D}$ falls in the allocation region of M. For such δ, a trade happens with probability at least $1/2$, with an expected profit of at least $1/4$. We refer to Figure 2b for a visualization.*

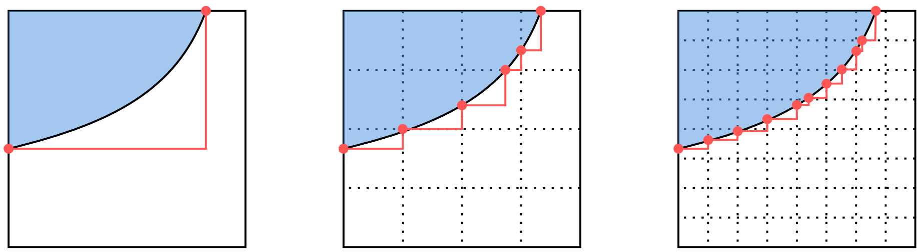
From Example 4.2, we know that past samples are not enough to estimate *at the same time* all mechanisms' expected performance. Note, the absence of uniform convergence is peculiar to our problem, and separates it to other mechanism design learning tasks such as, e.g., one-sided pricing [Kleinberg and Leighton 2003], revenue maximization in second price auctions with reserve [Cesa-Bianchi et al. 2015; Morgenstern and Roughgarden 2016], and gain-from-trade maximization with a single fixed price [Cesa-Bianchi et al. 2024a]. Moreover, from Example 4.1, we know we cannot focus on learning *only* mechanisms on a *finite and fixed* discretization! Di Gregorio et al. [2025] overcome this challenge by constructing a data and distribution dependent family of mechanisms which can be learned very fast (with $1/\varepsilon^2$ samples) and which well approximates the optimal mechanism.

A crucial role in the construction is played by the families $\mathcal{M}_\eta$ of η-simple mechanisms for $\eta \in (0, 1]$, i.e., such that the boundary of their allocation regions are determined by at most $O(1/\eta)$-axis parallel segments each (see Figure 3). In particular, for any mechanism $M \in \mathcal{M}$ there exists an η-simple mechanism in $\mathcal{M}_\eta$ which "mimics" its allocation region with $O(1/\eta)$ corners such that

$$\mathbb{E}\left[\mathsf{profit}(M, v)\right] \leq \mathbb{E}\left[\mathsf{profit}(M_\eta, v)\right] + O(\eta). \tag{2}$$

Focusing on the ε-simple mechanisms already yields some positive results, as a delicate but direct union-bound argument shows that approximately $1/\varepsilon^3$ samples are enough to estimate up to precision $O(\varepsilon)$ all the mechanisms in $\mathcal{M}_\varepsilon$.

To do better, Di Gregorio et al. [2025] resort to a delicate application of probabilistic chaining [Talagrand 2014], which allows for a more fine-grained control of the concentration argument. Denote with $H = \log 1/\varepsilon$, where ε is the target precision. Using an involved "good event" analysis, for any sample S of n i.i.d. valuations $v^1, v^2, \ldots, v^n$, Di Gregorio et al. [2025] shows the existence of a target sub-family $\mathcal{M}^\star$ of the ε-simple mechanisms which can be used to perform the chaining analysis. In



(a) Boundary of $\phi_0(M)$). (b) Boundary of $\phi_2(M)$. (c) Boundary of $\phi_3(M)$.

Fig. 3: Visualization of the chain construction.

particular, $\mathcal{M}^\star$ can be partitioned at geometrically decreasing "levels" of precisions $(1, 1/2, 1/4, \ldots, 2^{-H})$ ensuring that (i) there exists a target mechanism $M^\star \in \mathcal{M}^\star$ which is $O(\varepsilon)$ far from optimum, (ii) each level of the partition has cardinality $\binom{2^{2h}}{2^h}$, and (iii) the following L^1-net property is respected: for any $M \in \mathcal{M}^\star$ and any precision 2^{-h} there exists a mechanism $\phi_h(M) \in \mathcal{M}^\star \cap \mathcal{M}_{2^{-h}}$ such that

$$\left| \frac{1}{n} \sum_{v^i \in S} \left(\mathsf{profit}(M, v^i) - \mathsf{profit}(\phi_h(M), v^i) \right) \right| \le O(2^{-h}). \tag{3}$$

These ingredients can then be combined in the standard probabilistic chaining analysis to get uniform convergence in the $\mathcal{M}^\star$ family with the desired sample complexity of $\tilde{O}(1/\varepsilon^2)$. In particular, the algorithm, which first examines the realized samples, then computes the empirical profit maximizers, and finally transforms them so that they lie in the $\mathcal{M}^\star$ family, achieves the desired learning outcome.

We conclude by mentioning that, interestingly, a similar *chaining-based* approach is adopted in Di Gregorio et al. [2026] against the smooth adversary. There, given the non-stationary nature of the adversary, the authors need to resort to the "algorithmic" version of chaining, inspired by Cesa-Bianchi et al. [2017].

5. GENERALIZATIONS AND RELATED MODELS

In this section, we present some models for bilateral trade—or related problems—that extend or vary the basic learning protocol.

5.1 Contextual Bilateral Trade

Gaucher et al. [2025] introduce a natural tweak: *contexts*. Namely, at each time step t, the agents are characterized by d-dimensional vectors (x_s^t, x_b^t), and their actual valuations at that time are given by the inner product with unknown but fixed weight vectors v_s and v_b.

In the contextual setting, bilateral trade takes the form of searching the space of possible weight vectors, using a specific type of queries: the learner posts two prices, and observes whether the agents accept or not. Taking inspiration from the Ellipsoid Pricing method for (one-sided) pricing [Cohen et al. 2020], Gaucher et al. [2025] shows that, for gain-from-trade, it is possible to achieve *logarithmic* regret with respect to a very powerful benchmark—the best omniscient pricing strategy—as long as two-bit feedback is available.

Following up on this work, Cosson et al. [2026] derive nearly minimax regret rates for both profit and gain-from-trade, leveraging the notion of Steiner potentials [Liu et al. 2021]. Surprisingly, it turns out that gain-from-trade maximization is “equivalent” to contextual search ($\tilde{\Theta}(d)$ regret)[1] while profit-maximization in bilateral trade and one-sided pricing yields the same rate ($\tilde{\Theta}(d \log\log T)$). When the learner has only access to one-bit feedback, the regret landscape becomes less clear.

Gaucher et al. [2025] also introduce the *noisy* version of the problem, where the agents' valuations are given by the context-weight inner product plus some i.i.d. noise. In this case, they prove a tight (up to polylog(T) terms and dimension dependence) $\Theta(T^{2/3})$ regret bound for gain-from-trade maximization. Note, given the stochastic nature of the data-generation model, the benchmark is less powerful: the standard fixed-price-in-hindsight adversary. [Coccia et al. 2026] generalizes the model further, by considering agents whose valuations are given by a fixed *Lipschitz function*. This problem is non-parametric in nature, but the authors are able to show a tight $\tilde{O}(T^{d/d-1})$ regret rate for gain-from-trade maximization, using only one-bit feedback, under the strong budget balance condition. This positive result is achieved via a geometric decomposition of the action space, in a way that is somehow reminiscent of *adaptive zooming* [Slivkins 2014] and *algorithmic chaining* [Cesa-Bianchi et al. 2017].

These works provide positive results for various notions of budget balance or suboptimal regret rates, but a clean answer remains elusive. For instance, we find the following question especially interesting: can we achieve sublinear regret when the valuations are linear and noisy, under the strongly budget-balanced condition, with only one-bit feedback?

5.2 Brokerage

Bolić et al. [2024] introduce the brokerage problem: a version of bilateral trade in which the identity of the seller and the buyer are not fixed in advance, but depend on the relative ordering of their valuations. Namely, there exists an underlying fixed distribution over $[0, 1]$ and, at each iteration, two independent samples are drawn from it; the larger of the two samples becomes the buyer, while the smaller becomes the seller. Thus, the two sides of the market are coupled by the same underlying distribution, and the agents' roles are endogenous.

In the full-feedback model, Bolić et al. [2024] prove a tight $\sqrt{T}$ regret rate for a general distribution. Under the σ-smoothness assumption, they further derive a surprising $1/\sigma \cdot \log T$ regret rate, which comes from a clever rewriting of the gain-from-trade objective in terms of the variance of the empirical mean of the samples observed. This rewriting allows the authors to exploit the exponential decay of such variance via standard concentration arguments. Bolić et al. [2024] also consider the weaker one-bit-feedback model, where they prove that no learning is achievable for general distributions, while smoothness allows for a $\sqrt{T/\sigma}$.

Following up on this initial work, Bachoc et al. [2025a] and Bachoc et al. [2025b] investigate the contextual version of the brokerage problem, from both the parametric and non-parametric perspectives. In another modification, Cesari and Colomboni [2025] consider the same brokerage model, with the difference that the learner is

[1] Here $\tilde{\Theta}$ hides terms polynomial in the dimension d

now interested in maximizing the number of trades (and no longer the resulting gain from trade).

A closely related mediated model is the repeated-mediated newsvendor problem in Bolić et al. [2025]. Here, the learner is again a mediator between two sides of a market, but the interaction is richer than single-unit bilateral trade: a supplier has a private production cost, a retailer has a private utility function, the mediator posts a unit trading price, and the retailer's response determines the quantity purchased. This model can be viewed as a newsvendor-type generalization of the same mediated-trade viewpoint underlying brokerage and bilateral trade.

Finally, Cesa-Bianchi et al. [2025] studies market making, where agents arrive sequentially, and the mechanism posts prices that determine whether an arriving agent trades as a seller or as a buyer. This is not a direct bilateral-trade generalization, but it is a natural market-microstructure counterpart of brokerage: the side of the trade is again determined endogenously by the relationship between the agent's valuation and the posted prices.

5.3 Fair Bilateral Trade

One drawback of the gain-from-trade metric is that it measures only the *increase* in social welfare and provides no information about each agent's utility. For instance, consider a seller with valuation $v_s = 0$, and a buyer with $v_b = 1$; whichever price the mechanism may post in the $[0, 1]$ interval results in a trade with an increase in social welfare of 1. However, some prices may disproportionately benefit one agent at the expense of the other. Posting $p = 1$ increases the seller's utility from 0 to 1, while the buyer's utility remains constant at 0. To address this shortcoming, Bachoc et al. [2024] introduce the notion of *fair* gain from trade. Formally, the fair-gain-from-trade FGFT under valuations v_s and v_b for posting price p is:

$$\mathsf{FGFT}(p, v_s, v_b) = \min\{(v_b - p)_+, (p - v_s)_+\},$$

where $(a)_+ = \max\{0, a\}$. With this notion, the best price with valuations $(0, 1)$ becomes $1/2$, which divides the increase in utility between the two agents equally.

Bachoc et al. [2024] investigates various feedback and data generation models for the online version of the fair bilateral trade problem. In particular, for full feedback and stochastic i. adversary they show that the minimax regret is $\tilde{\Theta}(\sqrt{T})$ in the stochastic i.i.d. setting, while for one-bit-feedback, they show that no regret is only achievable assuming independence of the seller and buyer valuations, in which case the minimax rate is $\tilde{\Theta}(T^{2/3})$.

Building on Bachoc et al. [2024], Bachoc et al. [2026] develop an axiomatic fairness framework for bilateral trade, leading to a Rawls-to-Nash family of fair-gain objectives obtained by aggregating the seller's and buyer's gains through nonpositive Hölder means. This class leads to a rather involved statistical problem: estimating these objectives from threshold feedback requires handling two-dimensional singular-kernel identities. They obtain tight (up to logarithmic factors) regret guarantees, as well as PAC guarantees that hold uniformly and simultaneously over the fairness parameter.

5.4 Query Complexity

In Castiglioni et al. [2026], the authors investigate a fundamental learning problem that turns out to be intimately related to bilateral trade. Their goal is to get a *uniform* estimate of the multi-dimensional CDF of a distribution on the d-dimensional cube, using only "pricing queries" [Leme et al. 2023]. A pricing query takes in input d prices $p_1, \dots, p_d$, draws a fresh sample X from the underlying distribution, and return Yes if and only if $X \in [0, p_1] \times \cdots \times [0, p_d]$ (or, alternatively, it answer yes, if each coordinate of X *accepts* its price). [Castiglioni et al. 2026] shows how to solve the problem using $1/\varepsilon^3 (\log 1/\varepsilon)^{O(d)}$ pricing queries, which nearly-matches the 1-dimension lower bound of $1/\varepsilon^3$ implicit in Kleinberg and Leighton [2003].

Let's get back to bilateral trade and consider the task of finding the *fixed-price* mechanism that maximizes profit given a certain valuation distribution. In the i.i.d. setting, the task becomes finding the pair of prices $p \leq q$ maximizing $\mathbb{E}\left[\mathsf{profit}(M_{(p,q)}, v)\right] = (q - p) \cdot \mathbb{P}\left(v_s \leq p, v_b \geq q\right)$, while receiving one-bit feedback if the type $\mathbb{1}_{\{v_s \leq p, v_b \geq q\}}$. This problem is very similar to the one studied in Castiglioni et al. [2026], so that $\tilde{\Theta}(1/\varepsilon^3)$ samples are necessary and sufficient to get a uniform ε-estimate of the profit function! This implies that a suitably tuned explore-then-commit algorithm yields a $\tilde{O}(T^{3/4})$ regret: use the algorithm in Castiglioni et al. [2026] for $\approx T^{3/4}$ time steps to construct an $T^{-1/4}$-uniform estimate of the CDF of the distribution underlying (v_s, v_b), and then commit to the best pair of prices on the estimate.

An interesting direction for future work is pinpointing the right minimax regret achievable. Indeed, the best known lower bound is the $\Omega(T^{2/3})$ impossibility result for single-dimensional pricing in Kleinberg and Leighton [2003]. Notably, solving the stochastic setting is only the first step, as no tight result is known for the adversarial setting (where, with a bit of work, one could recover the same $T^{3/4}$ rate via a Lipschitzness argument).

5.5 Private Bilateral Trade

Some of the learning algorithms presented so far have the undesirable property of being too dependent on a few data points. For instance, the $\sqrt{T}$ algorithm for profit maximization in Di Gregorio et al. [2025] heavily relies on the realized points to construct the simple mechanism to play. This dependency leads to algorithms that are fragile to perturbed samples and, crucially, may publicly reveal information about the observed valuations. To address this weakness, Di Gregorio et al. [2026] investigate the bilateral trade problem through the lens of differential privacy (DP) within the PAC-learning framework. The authors establish a foundational impossibility result, proving that achieving both DP and near-optimality is inherently unattainable under general valuation distributions for both profit and gain-from-trade maximization. To bypass this barrier, they consider σ-smooth distributions and deliver nearly tight sample complexity bounds, showing that efficiency can be privately maximized using $\tilde{O}(1/\alpha^2 + 1/\epsilon\alpha)$ samples, where α is the approximation parameter, while ε is the privacy one. For profit maximization they need $\tilde{O}(1/\epsilon\sigma\alpha^2)$ samples. Beyond these statistical guarantees, a technical highlight of the work is algorithmic: they overcome the challenge of sampling from an exponentially large mechanism space by reducing to online shortest paths and exploiting a sampling trick from Takimoto and Warmuth [2003].

5.6 Beyond Bilateral Trade: Two-Sided Markets

Bilateral trade focuses on the interactions of *two* agents: a single seller and a single buyer. It is then natural to ask what happens in more complex *two-sided* markets. While the notions of incentive compatibility, individual rationality, and budget balance are easily generalizable, the induced class of mechanisms gets richer and loses the simple structure described in Propositions 2.2 and 2.5. In particular, gain-from-trade maximization now becomes richer than simply finding the best fixed price. In the online learning setting, two-sided markets have been investigated in Babaioff et al. [2024], Lunghi et al. [2025], and Feng et al. [2026]. In particular, Babaioff et al. [2024] show that even in small markets of three agents (one seller and two buyers, where at most one buyer can purchase the good from the seller), the problem is hopeless in the stochastic i.i.d. setting. Conversely, making the assumption that all the agents' valuations are independent is enough to guarantee a sample complexity of $\approx 1/\varepsilon^2$ (which corresponds, in the online setting, to $\approx \sqrt{T}$). Lunghi et al. [2025] instead focuses on larger markets, featuring one seller and an arbitrary number of buyers, but only restricts their attention to a simple family of mechanisms: second-price auction with reserve on the buyers' side and fixed price on the seller's side. For this family, they prove tight minimax results under various feedback and data generation models. Finally, Feng et al. [2026] extends the positive results for contextual bilateral trade to a more complex setting in which a single seller interacts with multiple buyers. Given these (partial) results, the study of two-sided markets in the online learning setting remains an exciting open direction to pursue.

5.7 Beyond DSIC mechanisms: Approximating the First Best

So far, the machine learning literature has mainly focused on studying the DSIC version of the problem, where incentive compatibility is ensured *regardless* of the other agent's behavior. Motivated by the fact that no DSIC and IR mechanism ensures a constant factor of gain-from-trade with budget balance [Blumrosen and Mizrahi 2016], Deng et al. [2025] investigates the learnability of the seller- and buyer-pricing mechanisms [Deng et al. 2022]. Indeed, it is known that uniformly randomizing between these two mechanisms (which enforce incentive compatibility only in expectation) yields a constant factor approximation of the so-called first best (i.e., $(v_b - v_s)_+$, see also discussion in Section 2.3).

Deng et al. [2025] considers the situation in which agents' valuations are drawn independently from unknown distributions. One of the two agents then observes its realized private valuation and is tasked with proposing a price to the other, using *only* sample access to the other agent's distribution. They investigate various sample-based pricing strategies and observe that, under some conditions, the resulting mechanisms yield, in expectation, a constant-factor approximation to the first best.

While Deng et al. [2025] adopts a different perspective than the one studied in the rest of the survey (there is "offline" sample access and the learner is one of the agents, and not the intermediary), it points to an exciting new research direction, namely investigating bilateral trade in the online learning framework *beyond* the DSIC + IR world, which is now fairly well understood. As a final (and related) research opportunity, we note that all previous work has made the simplifying assumption that an intermediary plays with a fresh pair of agents at each time step. This "myopic" version of incentive compatibility—incentives are measured only at each time step, without considering the possibility that agents might strategize about future time steps—is natural in some applications (e.g., a large population of "hasty" agents) and has allowed for a comprehensive (and satisfying) theoretical study. It is, however, natural to investigate what happens in a non-myopic world, where the agents are, for instance, no-regret learners (à la Braverman et al. [2018]); this problem is still wide open and is a natural next step in the study of bilateral trade.

Acknowledgment

The authors would like to thank Matteo Castiglioni, Tom Cesari, and Roberto Colomboni for their valuable feedback on preliminary versions of the survey.

REFERENCES

Awerbuch, B. and Mansour, Y. 2003. Adapting to a reliable network path. In *PODC*. ACM, 360–367.

Azar, Y., Fiat, A., and Fusco, F. 2024. An α-regret analysis of adversarial bilateral trade. *Artif. Intell. 337*, 104231.

Babaioff, M., Frey, A., and Nisan, N. 2024. Learning to maximize gains from trade in small markets. In *EC*. ACM, 195.

Bacchiocchi, F., Castiglioni, M., Colomboni, R., and Marchesi, A. 2025. Online bilateral trade with minimal feedback: Don't waste seller's time. In *NeurIPS*. Curran Associates, Inc.

Bachoc, F., Cesa-Bianchi, N., Cesari, T., and Colomboni, R. 2024. Fair online bilateral trade. In *NeurIPS*.

Bachoc, F., Cesari, T., and Colomboni, R. 2025a. A parametric contextual online learning theory of brokerage. In *ICML*. Proceedings of Machine Learning Research. PMLR / OpenReview.net.

Bachoc, F., Cesari, T., and Colomboni, R. 2025b. A tight regret analysis of non-parametric repeated contextual brokerage. In *AISTATS*. Proceedings of Machine Learning Research. PMLR, 2836–2844.

Bachoc, F., Colomboni, R., and Kaufmann, E. 2026. Repeated bilateral trade: The quest for fairness. *CoRR abs/2606.15369*.

Bernasconi, M., Castiglioni, M., Celli, A., and Fusco, F. 2024a. Bandits with replenishable knapsacks: the best of both worlds. In *ICLR*. OpenReview.net.

Bernasconi, M., Castiglioni, M., Celli, A., and Fusco, F. 2024b. No-regret learning in bilateral trade via global budget balance. In *STOC*. ACM, 247–258.

Blumrosen, L. and Dobzinski, S. 2021. (almost) efficient mechanisms for bilateral trading. *Games Econ. Behav. 130*, 369–383.

Blumrosen, L. and Mizrahi, Y. 2016. Approximating gains-from-trade in bilateral trading. In *WINE*. Lecture Notes in Computer Science. Springer, 400–413.

Bolić, N., Cesari, T., and Colomboni, R. 2024. An online learning theory of brokerage. In *AAMAS*. International Foundation for Autonomous Agents and Multiagent Systems / ACM, 216–224.

Bolić, N., Cesari, T., Colomboni, R., and Paravalos, C. 2025. Online Learning in the Repeated Mediated Newsvendor Problem. In *NeurIPS*.

Braverman, M., Mao, J., Schneider, J., and Weinberg, S. M. 2018. Selling to a no-regret buyer. In *EC*. ACM, 523–538.

Cai, Y. and Wu, J. 2023. On the optimal fixed-price mechanism in bilateral trade. In *STOC*. ACM, 737–750.

Castiglioni, M., Lunghi, A., and Marchesi, A. 2026. The sample complexity of uniform approximation for multi-dimensional cdfs and fixed-price mechanisms. *To appear at STOC 2026, preprint on the arXiv, CoRR abs/2602.10868*.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Foscari, L., and Pathak, V. 2025. Market making without regret. In *COLT*. Proceedings of Machine Learning Research. PMLR, 799–837.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Fusco, F., and Leonardi, S. 2024a. Bilateral trade: A regret minimization perspective. *Math. Oper. Res. 49,* 1, 171–203.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Fusco, F., and Leonardi, S. 2024b. Regret analysis of bilateral trade with a smoothed adversary. *J. Mach. Learn. Res. 25*, 234:1–234:36.

Cesa-Bianchi, N., Cesari, T. R., Colomboni, R., Fusco, F., and Leonardi, S. 2021. A regret analysis of bilateral trade. In *EC*. ACM, 289–309.

Cesa-Bianchi, N., Conconi, A., and Gentile, C. 2004. On the generalization ability of on-line learning algorithms. *IEEE Trans. Inf. Theory 50,* 9, 2050–2057.

Cesa-Bianchi, N., Gaillard, P., Gentile, C., and Gerchinovitz, S. 2017. Algorithmic chaining and the role of partial feedback in online nonparametric learning. In *Conference on Learning Theory*. PMLR, 465–481.

Cesa-Bianchi, N., Gentile, C., and Mansour, Y. 2015. Regret minimization for reserve prices in second-price auctions. *IEEE Trans. Inf. Theory 61,* 1, 549–564.

Cesa-Bianchi, N. and Lugosi, G. 2006. *Prediction, learning, and games*. Cambridge university press.

Cesari, T. and Colomboni, R. 2025. An online learning theory of trading-volume maximization. In *ICLR*. OpenReview.net.

Chen, H., Jin, Y., Lu, P., and Zhang, C. 2025. Tight regret bounds for fixed-price bilateral trade. *CoRR abs/2504.04349*.

Coccia, E., Bernasconi, M., and Celli, A. 2026. Nonparametric contextual online bilateral trade. In *ICLR*. OpenReview.net.

Cohen, M. C., Lobel, I., and Leme, R. P. 2020. Feature-based dynamic pricing. *Manag. Sci. 66,* 11, 4921–4943.

Cosson, R., Fusco, F., Gupta, A., Leonardi, S., Leme, R. P., and Russo, M. 2026. Contextual online bilateral trade. *To appear at EC'26, preprint on the arXiv CoRR abs/2602.12903*.

Deng, Y., Mao, J., Sivan, B., and Wang, K. 2022. Approximately efficient bilateral trade. In *STOC*. ACM, 718–721.

Deng, Y., Mao, J., Sivan, B., Wang, K., and Wu, J. 2025. Approximately efficient bilateral trade with samples. In *EC*. ACM, 206–223.

Di Gregorio, S., Dütting, P., Fusco, F., and Schwiegelshohn, C. 2025. Nearly tight regret bounds for profit maximization in bilateral trade. In *FOCS*. IEEE, 1570–1594.

Di Gregorio, S., Dütting, P., Fusco, F., and Schwiegelshohn, C. 2026. Profit maximization in bilateral trade against a smooth adversary. *CoRR abs/2605.12664*.

Di Gregorio, S., Fusco, F., Leonardi, S., and Schwiegelshohn, C. 2026. Private learning in bilateral trade. *CoRR abs/2606.02050*.

Dobzinski, S., Eden, A., Goldner, K., Shaulker, A., and Tsilivis, T. 2025. Bilateral trade with interdependent values: Information vs. approximation. In *EC*. ACM, 641–665.

Dütting, P., Fusco, F., Lazos, P., Leonardi, S., and Reiffenhäuser, R. 2026. Efficient two-sided markets with limited information. *SIAM J. Comput. 55,* 1, 65–92.

Fei, Y. 2022. Improved approximation to first-best gains-from-trade. In *WINE*. Lecture Notes in Computer Science. Springer, 204–218.

Feng, Y., Ma, M., Peng, B., and Wan, Z. 2026. Searching for optimal prices in two-sided markets. *To appear at EC'26, preprint on the arXiv CoRR abs/2602.11691*.

Gaucher, S., Bernasconi, M., Castiglioni, M., Celli, A., and Perchet, V. 2025. Feature-based online bilateral trade. In *ICLR*. OpenReview.net.

Hagerty, K. M. and Rogerson, W. P. 1987. Robust trading mechanisms. *Journal of Economic Theory 42,* 1, 94–107.

Haghtalab, N., Roughgarden, T., and Shetty, A. 2020. Smoothed analysis of online and differentially private learning. In *NeurIPS*.

Hajiaghayi, I., Hajiaghayi, M., Peng, G., and Shin, S. 2025. Gains-from-trade in bilateral trade with a broker. In *SODA*. SIAM, 4827–4860.

Jin, Y. 2026. Tight regret bounds for bilateral trade under semi feedback. *CoRR abs/2601.16412*.

Kang, Z. Y., Pernice, F., and Vondrák, J. 2022. Fixed-price approximations in bilateral trade. In *SODA*. SIAM, 2964–2985.

Kleinberg, R., Slivkins, A., and Upfal, E. 2019. Bandits and experts in metric spaces. *J. ACM 66,* 4, 30:1–30:77.

Kleinberg, R. D. and Leighton, F. T. 2003. The value of knowing a demand curve: Bounds on regret for online posted-price auctions. In *FOCS*. IEEE Computer Society, 594–605.

Kumar, R. and Kleinberg, R. 2022. Non-monotonic resource utilization in the bandits with knapsacks problem. In *NeurIPS*.

Leme, R. P., Sivan, B., Teng, Y., and Worah, P. 2023. Pricing query complexity of revenue maximization. In *SODA*. SIAM, 399–415.

Liu, A., Leme, R. P., and Schneider, J. 2021. Optimal contextual pricing and extensions. In *SODA*. SIAM, 1059–1078.

Liu, Z., Ren, Z., and Wang, Z. 2023. Improved approximation ratios of fixed-price mechanisms in bilateral trades. In *STOC*. ACM, 751–760.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2025. Online two-sided markets: Many buyers enhance learning. *CoRR abs/2503.01529*.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2026a. Better regret rates in bilateral trade via sublinear budget violation. In *SODA*. SIAM, 6494–6536.

Lunghi, A., Castiglioni, M., and Marchesi, A. 2026b. Regret minimization in bilateral trade with perturbed markets. *CoRR abs/2605.10475*.

Morgenstern, J. and Roughgarden, T. 2016. Learning simple auctions. In *COLT*. JMLR Workshop and Conference Proceedings, vol. 49. JMLR.org, 1298–1318.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of operations research 6,* 1, 58–73.

Myerson, R. B. and Satterthwaite, M. A. 1983. Efficient mechanisms for bilateral trading. *Journal of economic theory 29,* 2, 265–281.

Roughgarden, T. 2020. *Beyond the Worst-Case Analysis of Algorithms*. Cambridge University Press, Cambridge, UK.

Slivkins, A. 2014. Contextual bandits with similarity information. *Journal of Machine Learning Research 15*, 2533–2568.

Slivkins, A., Zhou, X., Sankararaman, K. A., and Foster, D. J. 2024. Contextual bandits with packing and covering constraints: A modular lagrangian approach via regression. *J. Mach. Learn. Res. 25*, 394:1–394:37.

Takimoto, E. and Warmuth, M. K. 2003. Path kernels and multiplicative updates. *J. Mach. Learn. Res. 4*, 773–818.

Talagrand, M. 2014. *Upper and lower bounds for stochastic processes*. Vol. 60. Springer, Berlin, Heidelberg, Germany.

Vickrey, W. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance 16,* 1, 8–37.