

Gap-Based Learning in Multi-Unit Markets

Rigel Galgana, Negin Golrezaei, Yifeng Teng

MIT Operations Research, MIT Sloan School of Management, Google Research

Authors are encouraged to submit new papers to INFORMS journals by means of a style file template, which includes the journal title. However, use of a template does not certify that the paper has been accepted for publication in the named journal. INFORMS journal templates are for the exclusive purpose of submitting to an INFORMS journal and are not intended to be a true representation of the article's final published form. Use of this template to distribute papers in print or online or to submit papers to another non-INFORM publication is prohibited.

Abstract.

Theoretical predictions in markets and auctions can diverge substantially from observed outcomes, such as unexpectedly low revenue in multi-unit pollution-license auctions. Understanding competitive outcomes therefore requires studying both static solution concepts and the behavioral dynamics through which they emerge. No-regret learning and mean-based algorithms (MBAs), together with their connections to coarse correlated equilibria (CCEs) and rationalizability, have partially bridged these static and dynamic perspectives. However, no-regret learning imposes strong performance requirements, while MBAs place relatively little structure on behavior. We introduce gap-based algorithms (GBAs), an intermediate class that prunes actions with large performance gaps while continuing to explore among actions with small gaps. We show that these two properties yield distinct restrictions on equilibrium convergence. First, large-gap pruning requires every action in each player's limiting support to be a best response, refining CCEs to a class we call best-response-supported CCEs (BR-CCEs). Second, small-gap exploration implies that substitutes—best responses outside an equilibrium's limiting support—can destabilize convergence to that equilibrium.

We apply our analysis to multi-unit auctions, specifically pay-as-bid (PAB) and uniform-price auctions (UPAs). We show that BR-CCEs more accurately capture the welfare and revenue outcomes of learning dynamics in both PAB and UPA. Furthermore, when linearly suboptimal opponent actions are pruned sufficiently quickly, we show that our small gap exploration property prohibits convergence towards positive-revenue pure Nash equilibria (PNEs) in UPA, helping explain the low revenues observed in pollution-license auctions.

Key words: gap-based learning, learning in games; auctions; equilibrium selection; multi-unit auctions

1. Introduction

Learning algorithms are now routinely used in repeated market environments such as online advertising auctions, pricing platforms, and automated procurement. In these markets, agents repeatedly update prices or bids from historical feedback, and the long-run outcome is determined

not just by the underlying static game but by the way the agents’ algorithms treat actions that have been nearly optimal versus consistently sub-optimal. This is especially important in markets with flat payoff regions—i.e., regions where many actions yield identical payoffs. For example, a losing bid in a second-price auction may generate zero payoff for its bidder, but it still sets the winner’s price. The same holds true for its multi-unit variant—the uniform-price auction (UPA), formally defined in Section 6.1.

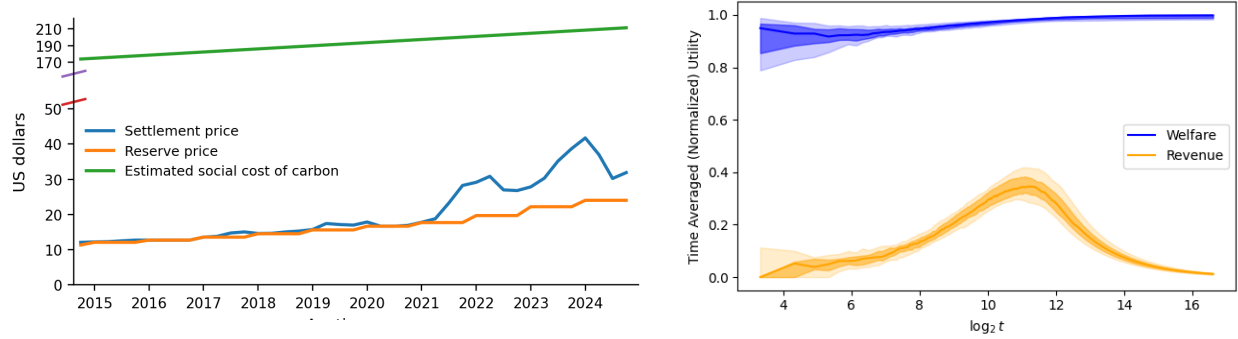


Figure 1 Uniform-price auction outcomes in practice and simulation. In the left figure, California’s pollution allowance license prices remain close to its regulatory floor [45], far below the estimated social cost of carbon [93]. In the right figure, the simulated learning dynamics of the uniform-price auction (Section E.2) consistently converge to 0 revenue, suggesting implicit algorithmic collusion to drive prices down.

One application is the California–Québec carbon market, where firms purchase emissions permits through uniform-price auctions and the proceeds fund clean up programs [15, 44]. Allowance prices have frequently remained near the regulatory floor (Figure 1) [15]. The 2020 auction prices of US\$16.68–17.87 per ton of CO₂ were far below the U.S. Environmental Protection Agency’s 2023 central social cost estimate of US\$190 per ton for 2020 emissions (2020 dollars, 2% near-term discount rate) [45, 93]. These poor auction revenues result in low funding for clean up efforts [15]. To explain this, academic studies provide evidence of low revenues from the standpoint of *implicit algorithmic collusion* [32, 55]. Indeed, our UPA simulations exhibit a related pattern: revenue approaches consistently approaches 0 as we see in Figure 1.

Understanding the reason behind such algorithmic collusion requires a model of how learning algorithms interact collectively. In the context of auctions where payoff functions are often flat—i.e., a losing bid of b versus b' does not affect one’s allocation nor payment—it will also be important to also understand how these algorithms behave when there are many actions with similar rewards. However, the standard algorithmic classes considered in the literature—mean-based algorithms

(MBAs), no-regret algorithms, and Follow-the-Regularized-Leader (FTRL) algorithms, formally defined in Definition 5—are either too structured or too unstructured to select meaningful outcomes [22, 36] or be representative of typical human decision-making behavior [38]. At a high level:

1. **No-regret algorithms, including standard no-regret FTRL parameterizations**, have been studied extensively because their empirical distributions approach a relaxed version of Nash equilibria (NE, Definition 1) called coarse correlated equilibria (CCE, Definition 6); when the empirical distributions converge in the time-average sense of Definition 3, their limits lie in CCE. That being said, typical human decision making often deviates from no-regret dynamics [74]. Experimentally, simply taking worst-case instances—as measured by the price of anarchy (PoA, Section 4.1)—over CCEs also fails to accurately predict the outcome of learning dynamics as we see in the overly pessimistic revenue and welfare estimates (Table 3) for the pay-as-bid (PAB) multi-unit auction formally defined in Section 6.1.
2. **MBAs** require that highly sub-optimal actions be chosen rarely [11]. This property is strong enough to prune actions outside the correlated-rationalizable action set formally defined in Proposition 2 [9], but it places no lower bound on near-optimal actions—in fact, the hindsight-best action may be assigned 0 probability, even when the worst actions are chosen with positive probability. Consequently, MBA dynamics can yield unnatural outcomes, such as the cyclical bidding behavior in the first-price auction (FPA) in [22, Example 7], whose time average converges to a non-NE profile.

The unintended effects of implicit algorithmic collusion cannot be identified simply with standard equilibrium analysis. For example, UPAs admit many positive-revenue equilibria [66] yet the revenue often remains near 0 (Figure 1). We therefore study how interactions among learning algorithms—for which we refer to as *learning dynamics*—gravitate towards certain outcomes over others. This lead to our main research question:

Can we better explain and predict the outcomes of algorithmic collusion in markets by analyzing the nature of the learning dynamics themselves?

1.1. Contributions

Contribution 1: Relation to existing algorithm classes. We show that gap-based algorithms (GBAs, Definition 4) serve as an intermediary class of learning algorithms between MBAs and FTRL in Proposition 1. At a high level, a GBA assigns each action a probability that decreases with its cumulative dynamic gap (Equation (4)): large-gap actions have vanishing probability, just like

MBAs. What distinguishes GBAs from MBAs is that small-gap actions retain probability bounded away from zero, similar to FTRL algorithms. Its sensitivity determines how quickly probability decreases with the gap, while its tolerance determines how quickly the gap must grow before pruning begins. These two restrictions lead to the large-gap and small-gap results below, which we use to characterize *non-convergence* to certain equilibria.

Contribution 2: Large-gap pruning. Our second main contribution is showing that the large-gap pruning property in Definition 4 restricts not only what actions can appear in each player’s support of a limiting distribution, but also that all such actions must best respond to the limiting opponent distribution. Theorem 1 formalizes this requirement as support optimality. Coupling this property with no regret restricts every empirical limit to the best-response-supported CCE (BR-CCE) set formally defined in Definition 6 (Proposition 3). This set inclusion is not purely cosmetic and has non-trivial implications for welfare guarantees in multi-unit auctions and their corresponding learning dynamics. In particular, for small instances of PAB, we show that this BR-CCE refinement improves the PoA over that of CCEs (Theorem 2), implying that no-regret GBA dynamics achieve a higher welfare guarantee when they converge compared to when they do not. For larger instances of PAB, we experimentally demonstrate that the GBA dynamics actually reflect this improved welfare—and even revenue—of BR-CCE over CCE and its CR-restricted version CCE_{CR} (Definition 6) in Section 6.2.

Contribution 3: Small-gap exploration. Our third contribution concerns unsupported best responses, which we call substitutes in Definition 2. Here, we call an opponent profile a *stay* profile (Definition 7) when some action in the limiting support strictly outperforms the substitute. Theorem 3 shows that time-average convergence to a distribution excluding this substitute requires the cumulative stay frequency to grow rapidly. If it does not, the substitute will still be explored often (Small-gap exploration, Definition 4). The forgiving condition in Definition 8 gives a direct way to show that stay profiles occur too infrequently. In particular, Theorem 4 assumes that every stay profile contains an opponent action with positive static gap (Equation (2)), which we show accumulates linear dynamic gap (Equation (4)) in Proposition 12. The forgiveness algorithmic condition guarantees that the relevant opponent action has cumulative frequency within player n ’s tolerance, ruling out convergence to such a limit.

Contribution 4: Applications. Our main application is the analysis of the dynamic welfare and revenue of multi-unit auctions. First, using the large-gap pruning property, we compute BR-CCE experiments for both PAB and UPA. These experiments show that BR-CCE is especially predictive

for PAB (Table 3). However, for UPA, our simulated GBA dynamics consistently converge to 0 revenue whereas BR-CCE still admits positive revenue (Table 3). In other words, large-gap pruning alone cannot explain the low revenue of UPA. To address this, we turn to small-gap exploration. We give two intermediary results—bilateral-trade (Proposition 4) and FPA (Corollary 2)—as warm-ups to illustrate small-gap exploration through forgiving dynamics before giving our main UPA theorem. In particular, our main UPA theorem (Theorem 5) shows that forgiving GBA dynamics cannot converge to positive-revenue pure Nash equilibria (PNEs, Definition 1), which helps to explain the 0 revenue convergence in our simulations.

1.2. Related work

We organize the related work around three themes. Section 1.2.1 discusses learning-based equilibrium refinements and stability. Section 1.2.2 discusses learning in auctions, pricing, and competition; Section 1.2.3 discusses multi-unit auction formats and market applications.

1.2.1. Equilibrium refinements and stability The works in this subsection are closest in spirit to ours: they ask how learning rules refine the broad equilibrium predictions obtained from standard regret guarantees. Standard external no-regret learning famously converges to the CCE set [17]. However, CCE can be too broad as an outcome prediction in market games with many substitutes (Definition 2). A growing literature therefore studies how the learning rule refines the set of possible supports. In Table 1, we describe the most relevant dynamics and their corresponding solution concepts. First, [36, 95] show that strict PNEs (Definition 2), if they exist, are the uniquely stable and attracting limit points of FTRL. In [9], the authors show that MBAs [11] prune actions outside the correlated rationalizable (CR) action set (Proposition 2). In particular, the CR restriction in CCE_{CR} —a solution concept closely related to the co-undominated CCE defined in [55]—requires each retained action to be weakly optimal against some joint distribution over the opponents’ retained actions (Proposition 2). Coupling this action-space restriction with CCE gives CCE_{CR} (Definition 6). Similar action-space restrictions also appear for dynamics beyond MBAs and GBAs, including gradient-descent dynamics and correlated-belief equilibrium [1, 2, 92]. Our large-gap result is stricter in a different direction: every action in the limiting support must be optimal against the actual limiting opponent distribution, not merely against *some* arbitrary distribution supported on CR actions. We use this tighter restriction to show that convergent dynamics yield stronger welfare guarantees than non-convergent ones in repeated PAB auctions.

Literature / class	Synopsis	Comments on Solution Concept
No-regret	No (external) regret algorithms guarantee that strategies incur vanishing sub-optimality relative to the hindsight best fixed action. If the empirical distribution converges, it must do so to a CCE [17].	CCE is a broad benchmark: it can support actions outside the CR set, and supported actions need not be optimal against the limiting opponent distribution. This can make CCE-based PoA overly pessimistic [26, 51, 90].
MBAs [11]	Prior MBA analyses show that actions outside the CR action set are pruned [9].	Theorem 1 strengthens this restriction: every empirical MBA limit lies in $\mathcal{P}^*$, and every no-regret MBA limit lies in BR-CCE.
FTRL [36, 95]	An equilibrium is simultaneously stable and attracting if and only if it is a strict PNE (Definition 2).	PNEs, let alone strict PNEs, need not exist. Some equilibria may also be attracting but not stable—arguably a more natural criterion.
GBAs (This work)	Large-gap pruning guarantees that all limiting supported actions are best responses, while small-gap exploration shows certain unsupported best responses cannot be pruned at a particular equilibrium.	Coupled with no-regret, the GBA empirical limit converges to the class of BR-CCEs (Definition 6). Theorem 4 gives an additional non-convergence condition for substitutes.

Table 1 A compact comparison of the solution concepts and convergence predictions most closely related to our results. CCE, CCE_{CR} , and BR-CCE are defined in Definition 6; CR support and $\mathcal{P}^*$ in Proposition 2; and strict PNEs in Definition 2.

1.2.2. Learning in auctions, pricing, and competition A broad literature studies auctions and market games through repeated learning dynamics. In FPAs, classical work analyzes equilibrium structure [7, 57, 58, 67], while more recent work studies bidder learning, MBA rules, regret, and computation [6, 22, 47, 48, 76, 97]. Discrete equilibrium computation and complexity in FPAs are studied in [19, 29, 30]. In second-price and reserve-price environments, the literature studies revenue optimization, contextual or semi-bandit feedback, and dynamic incentive-aware learning [39, 40, 69, 70, 73, 98]. Position auctions and sponsored-search markets are studied in [5, 23, 37, 94]. Bilateral trade has also been studied from a regret-minimization perspective, with attention to gains from trade and budget-balance constraints [8, 16, 64]. Multi-unit auctions form another branch of this learning literature and connect the bidder-side learning papers above to mechanism-level welfare and revenue comparisons [12, 32, 41, 42, 77, 78].

Learning is also used to study pricing, competition, and other economic games. MBA learning was introduced in a dynamic selling problem [11], and related pruning or convergence ideas appear in pricing, principal-agent, and persuasion settings [28, 61]. Bertrand and Cournot competition provide classical models of price and quantity competition [71, 85]. In learning-based versions of these games, outcomes can depend strongly on the learning rule and the market primitives [9, 33, 46]. More generally, multi-agent learning papers develop guarantees in structured games through predictable-sequence methods, robust fast convergence, and optimistic learning [18, 31, 79, 89].

Related last-iterate and regret guarantees appear in [3, 20, 38, 59]. Our paper contributes to this literature by studying the market outcome selected by a class of GBA rules rather than by designing a learning algorithm for one bidder or bounding regret in isolation.

1.2.3. Multi-unit auctions and market applications Multi-unit auctions are used whenever multiple homogeneous or closely related units are allocated at once, and the payment rule often determines whether bidders pay their own bids, as in PAB, or a common market-clearing price, as in UPA. Treasury auctions are a canonical application, and the literature has long compared UPA and PAB formats in that setting [10, 35]. Electricity markets provide another major application: market operators often need to clear many units of electricity while balancing market power, demand uncertainty, and operational constraints [24, 25, 83, 91]. Emissions-license and carbon-allowance markets also rely on multi-unit auction formats [43, 63, 75, 84]. Auction revenues in these markets fund environmental, energy-efficiency, renewable-energy, public-health, and household-assistance programs [50, 80, 81, 86]. In these markets, low prices can be economically important because allowance prices may sit below estimates of marginal climate damages or the social cost of carbon [13, 14, 53, 54, 93]. Related multi-unit and combinatorial auction ideas also appear in spectrum auctions, where simultaneous ascending and telecom-license auctions are central examples [52, 68], and in procurement and transportation-service markets, where auction rules must handle variable quantities, service attributes, and market-clearing constraints [4, 56, 82].

Recent work studies multi-unit market design through bidder-side learning, equilibrium computation, and market-clearing algorithms. Repeated PAB auctions are studied in [32]. Related work compares PAB and UPA under regret minimization, improves learning rates in repeated UPA, and studies safe or budget-constrained bidder learning [12, 41, 42, 77, 78]. Static and behavioral analyses of UPA and repeated multi-unit auctions include NE computation and AI or machine behavior in repeated auction environments [27, 60, 62]. Other market-clearing work studies simple pricing and constrained Walrasian equilibria [34, 65], as well as smart-grid, coupled electricity-carbon, and privacy-preserving clearing systems [49, 72, 87, 88]. These papers largely ask how bidders can learn, how equilibria behave, or how multi-unit markets can be cleared. Our focus is different: we ask which outcomes are selected when all bidders use interacting learning dynamics, and we show that PAB and UPA differ because substitutes have different price effects under the two payment rules.

2. Preliminaries

Notation. For an integer k , write $[k] \doteq \{1, \dots, k\}$. We use bold symbols for vectors and profiles, e.g., $\mathbf{a} = (a^1, \dots, a^N)$, and write $\mathbf{a}^{-n}$ for all components except player n . For a finite set S , let $|S|$ denote its cardinality and $\Delta(S)$ the set of probability distributions over S ; let $\mathbf{1}\{E\}$ denote an indicator; and write $[x]^+ \doteq \max\{x, 0\}$. For any distribution $\mathbf{p}$ over a finite product action space, $\mathbf{p}^n$ denotes the marginal distribution of player n , $\mathbf{p}^{-n}$ denotes the joint marginal distribution of player n 's opponents, and $\text{Supp}(\mathbf{p}^{-n}) \doteq \{\mathbf{a}^{-n} : \mathbf{p}^{-n}(\mathbf{a}^{-n}) > 0\}$ denotes the strict support of this opponent joint distribution. All asymptotic comparisons are as $t \rightarrow \infty$, and constants may depend on the finite one-shot game, the action spaces, and fixed algorithm parameters, but never on t . For nonnegative sequences x_t, y_t , write $x_t \lesssim y_t$ if $x_t \leq C y_t$ eventually for some finite constant C , write $x_t \ll y_t$ if $x_t/y_t \rightarrow 0$, write $x_t \gtrsim y_t$ and $x_t \gg y_t$ for the reverse relations, and write $x_t \asymp y_t$ if both $x_t \lesssim y_t$ and $y_t \lesssim x_t$.

Static game setting. We consider a finite normal-form game with players $n \in [N]$, finite action sets $\mathcal{A}^n$, joint action space $\mathcal{A} \doteq \prod_n \mathcal{A}^n$, and payoffs $u^n(a^n, \mathbf{a}^{-n})$. With abuse of notation, $\mathbf{p}$ may denote either a product mixed profile or a correlated joint distribution over $\mathcal{A}$; when we say that $\mathbf{p}$ is an NE, we mean a product distribution. For any joint distribution $\mathbf{p} \in \Delta(\mathcal{A})$, write

$$u^n(a, \mathbf{p}^{-n}) \doteq \sum_{\mathbf{a}^{-n}} u^n(a, \mathbf{a}^{-n}) \mathbf{p}^{-n}(\mathbf{a}^{-n}), \quad u^n(\mathbf{p}) \doteq \sum_{\mathbf{a}} u^n(\mathbf{a}) \mathbf{p}(\mathbf{a}). \quad (1)$$

We define the *static gap* of action $a \in \mathcal{A}^n$ against the opponent joint distribution $\mathbf{p}^{-n}$ as

$$\delta^n(a; \mathbf{p}) \doteq \max_{b \in \mathcal{A}^n} u^n(b, \mathbf{p}^{-n}) - u^n(a, \mathbf{p}^{-n}). \quad (2)$$

Thus a is a best response to $\mathbf{p}^{-n}$ if and only if $\delta^n(a; \mathbf{p}) = 0$, and it is strictly suboptimal with respect to $\mathbf{p}$ if $\delta^n(a; \mathbf{p}) > 0$. With some abuse of notation, we define $\delta^n(a; \mathbf{a})$ as the static gap of a w.r.t. the pure profile $\mathbf{a}$.

DEFINITION 1 (NASH EQUILIBRIA). A pure action profile $\mathbf{a}$ is a pure Nash equilibrium (PNE) if $\delta^n(a^n; \mathbf{a}) = 0$ for every player n . A mixed Nash equilibrium (MNE) is a product distribution $\mathbf{p} = \prod_{n \in [N]} \mathbf{p}^n$ such that every action in every player's support is a best response:

$$a \in \text{Supp}(\mathbf{p}^n) \implies \delta^n(a; \mathbf{p}) = 0 \quad \forall n \in [N], a \in \mathcal{A}^n.$$

Several of our later results concern non-convergence to certain weak NE. We therefore isolate the following subclass of NE and the associated notion of a substitute.

DEFINITION 2 (SUBSTITUTES AND WEAK NASH EQUILIBRIA). For any possibly correlated distribution $\mathbf{p}$, an action $a \notin \text{Supp}(\mathbf{p}^n)$ is a *substitute* for player n if it is a best response to $\mathbf{p}^{-n}$, i.e., $\delta^n(a; \mathbf{p}) = 0$. The substitute set is

$$\mathcal{S}^n(\mathbf{p}) \doteq \{a \notin \text{Supp}(\mathbf{p}^n) : \delta^n(a; \mathbf{p}) = 0\}.$$

An NE is *weak* if at least one player has a substitute. A PNE is *strict* if no player has a substitute, i.e., each player's action is its unique best response.

Repeated game setting. Now consider the repeated game. At round t , player n submits a mixed strategy $\mathbf{p}_t^n \in \Delta(\mathcal{A}^n)$, and $\mathbf{p}_t^{-n}$ denotes the distribution of the other players' mixed strategies. We work in the expected-feedback model, where agents can see the historical mixed strategy profiles of other agents.¹ For each player n and action a , define cumulative utility and dynamic gap by

$$U_t^n(a) \doteq U_0^n(a) + \sum_{\tau < t} u^n(a, \mathbf{p}_\tau^{-n}), \quad (3)$$

$$\Delta_t^n(a) \doteq \max_{a' \in \mathcal{A}^n} U_t^n(a') - U_t^n(a). \quad (4)$$

where initialization terms $U_0^n(a)$ may encode priors or previous play.² Define the empirical joint distribution of mixed play, its player- n marginal, and cumulative action frequency by

$$\bar{\mathbf{p}}_t(\mathbf{a}) \doteq \frac{1}{t} \sum_{\tau < t} \prod_{i \in [N]} p_\tau^i(a^i), \quad \bar{\mathbf{p}}_t^n(a) \doteq \sum_{\mathbf{a}: a^n = a} \bar{\mathbf{p}}_t(\mathbf{a}) = \frac{P_t^n(a)}{t}, \quad P_t^n(a) \doteq \sum_{\tau < t} p_\tau^n(a). \quad (5)$$

For a set $S \subseteq \prod_{i \neq n} \mathcal{A}^i$ of opponent profiles, define its cumulative frequency by

$$P_t^{-n}(S) \doteq \sum_{\tau < t} p_\tau^{-n}(S) = \sum_{\tau < t} \sum_{\mathbf{a}^{-n} \in S} \prod_{i \neq n} p_\tau^i(a^i). \quad (6)$$

We write $P_t^{-n}(\mathbf{a}^{-n})$ for $P_t^{-n}(\{\mathbf{a}^{-n}\})$. Thus $\bar{\mathbf{p}}_t^{-n}(S) = P_t^{-n}(S)/t$. With this, we can now define our main notion of convergence.

DEFINITION 3 (TIME-AVERAGE CONVERGENCE). The dynamics converge in time-average—i.e. *converge* for short—to a possibly correlated distribution $\mathbf{p}$ if $\lim_{t \rightarrow \infty} \bar{\mathbf{p}}_t = \mathbf{p}$.

The product inside $\bar{\mathbf{p}}_t$ reflects *decentralized* mixed play where in each round, players randomize independently, so the joint action profile distribution is the product of their mixed strategies.

¹ This expected-feedback assumption is mostly to simplify our analysis. We discuss in Appendix A.3 a high-probability realized-feedback analogue of the static-gap bounds used in our proofs.

² We note that the initialization bias is a constant, whereas our results pertain to asymptotics. While the initialization may change the eventual equilibrium converged to, it does not change our non-convergence criteria (Theorems 1 and 3).

2.1. Gap-Based Algorithms

Now, we define the titular class of algorithms that will be the main object of focus in our paper.

DEFINITION 4 (GAP-BASED ALGORITHM (GBA)). A GBA is parameterized by its *sensitivity* $\gamma_t^n : \mathbb{R}_+ \rightarrow [0, 1]$ —which controls how quickly large-gap actions are discarded—and its *tolerance exponent* $\alpha_n \in (0, 1)$ —which gives the tolerance order t^{α_n} . For every $a \in \mathcal{A}^n$, a GBA guarantees

$$p_t^n(a) \asymp \gamma_t^n(\Delta_t^n(a)), \quad (7)$$

where $\gamma_t^n(\Delta)$ is weakly decreasing in Δ . For every sub-optimality sequence $\Delta = (\Delta_t)_{t>0}$, the small-gap and large-gap conditions are

$$\Delta_t \gg t^{\alpha_n} \Rightarrow \gamma_t^n(\Delta_t) \ll 1. \quad (\text{Large Gap Pruning})$$

$$\Delta_t \lesssim t^{\alpha_n} \Rightarrow \gamma_t^n(\Delta_t) \asymp 1. \quad (\text{Small Gap Exploration})$$

To interpret these conditions, the large-gap condition states that whenever some action a has sub-optimality growing more *quickly* than t^{α_n} , then $p_t^n(a) \rightarrow 0$. This property is used in Appendix A.1 when translating the GBA large-gap condition into the MBA pruning bound. The small-gap condition guarantees that any suboptimality sequence at most order t^{α_n} must eventually have strictly positive probability. This exploration is *not* present in MBAs: actions with low sub-optimality are not guaranteed to have higher probability than those with large sub-optimality. This drives much of the unintuitive behavior seen in MBA dynamics [22].

3. Relationship with Other Algorithm Classes

The large-gap and small-gap conditions of GBAs resolve the aforementioned MBAs' probability non-monotonicity while relaxing the structure of FTRL algorithms. For completeness, we define these algorithm classes here:

DEFINITION 5 (STANDARD ALGORITHM CLASSES). A learning rule is an MBA if there exists $\beta_t \rightarrow 0$ such that $\Delta_t^n(a) \geq t\beta_t$ implies $p_t^n(a) \leq \beta_t$ for every player n and action a . It has *no external regret* if, for every player n :

$$\max_{a \in \mathcal{A}^n} U_t^n(a) - \sum_{\tau < t} u^n(\mathbf{p}_\tau) = o(t).$$

An FTRL rule, for a convex separable regularizer $\Psi^n(\mathbf{x}) = \sum_{a \in \mathcal{A}^n} \psi_a^n(x(a))$ (symmetric if every ψ_a^n is the same function) and scale η_t , chooses³

$$\mathbf{p}_t^n \in \arg \max_{\mathbf{x} \in \Delta(\mathcal{A}^n)} \left\{ \sum_{a \in \mathcal{A}^n} x(a) U_t^n(a) - \eta_t \Psi^n(\mathbf{x}) \right\}.$$

Having defined the closest algorithm classes, we next state their relationship to GBAs.

PROPOSITION 1 (Algorithm-class relationships). *GBAs and no-regret algorithms are incomparable (i.e., not subsets of one another), and GBAs form a strict subclass of MBAs. Moreover, every symmetric separable FTRL rule satisfying Proposition 8 is a GBA.*

In Appendix A.1, we state the formal FTRL requirements and prove these relationships. The following list summarizes the algorithms used in our examples and experiments (Section 6). Some are GBAs, while the others are included only for comparison. For each GBA example, we state its tolerance exponent and a valid sensitivity from Definition 4; we show the full definitions and proofs in Appendix A.2.

- **FTRL and Hedge.** For the symmetric separable FTRL rules in Proposition 8, the regularization is of order t^{α_n} , so the GBA tolerance exponent is α_n . With ψ denoting the regularizer in that proposition, a valid sensitivity is the unique $x \in (0, 1]$ satisfying $\psi'(1) - \psi'(x) = \Delta/t^{\alpha_n}$. For Hedge (entropy FTRL), its tolerance exponent is α_n and has exponential sensitivity:

$$\gamma_t^n(\Delta) = e^{-\Delta t^{-\alpha_n}}, \quad p_t^n(a) = \frac{\gamma_t^n(\Delta_t^n(a))}{\sum_{a' \in \mathcal{A}^n} \gamma_t^n(\Delta_t^n(a'))}.$$

- **Polynomial weights and Tsallis FTRL.** Polynomial weights and Tsallis FTRL with parameter $q_{ts} \in (0, 1)$ (defined in Appendix A.2), with tolerance exponents α_n , both have polynomial decays as a function of $\Delta t^{-\alpha_n}$ of order $\beta_n > 0$ and $1/(1 - q_{ts})$ respectively. Their sensitivities are

$$\gamma_t^n(\Delta) = (1 + \Delta t^{-\alpha_n})^{-\beta_n} \quad \text{and} \quad \gamma_t^n(\Delta) = (1 + (1 - q_{ts}) \Delta t^{-\alpha_n})^{-1/(1 - q_{ts})}$$

respectively. We note that while the two algorithms look similar with polynomial pruning rates, the Polynomial weights algorithm is a GBA but is not generally FTRL-based or no-regret unlike Tsallis FTRL.

³ For the FTRL rules that we classify as GBAs, Proposition 8 imposes a steepness condition that keeps every probability positive. Euclidean FTRL does not satisfy that condition and may assign zero probability; Appendix Sections A.1 and A.2 discuss this boundary case.

- **ϵ_t -buffered GBAs.** Given a base GBA with sensitivity $\gamma_{t,0}^n$ and tolerance exponent α_n , an ϵ_t -buffered GBA simply adds $\epsilon_t^n/|\mathcal{A}^n|$ uniform noise to all action probabilities:

$$p_t^n(a) = (1 - \epsilon_t^n)p_{t,0}^n(a) + \frac{\epsilon_t^n}{|\mathcal{A}^n|}, \quad \gamma_t^n(\Delta) = (1 - \epsilon_t^n)\gamma_{t,0}^n(\Delta) + \epsilon_t^n.$$

It retains tolerance exponent α_n .

- **Euclidean FTRL, Follow-the-Leader, and Gradient Descent.** Euclidean FTRL sets $\Psi^n(\mathbf{x}) = \frac{1}{2} \sum_a x(a)^2$. Follow-the-leader (FTL) chooses a cumulative-utility maximizer, $a_t^n \in \arg \max_{a \in \mathcal{A}^n} U_t^n(a)$ and puts probability one on a_t^n . Projected gradient descent (PGD), with step size $\eta_t > 0$, updates according to

$$\mathbf{p}_{t+1}^n = \Pi_{\Delta(\mathcal{A}^n)}(\mathbf{p}_t^n + \eta_t \mathbf{u}_t^n), \quad \mathbf{u}_t^n = (u^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n},$$

where $\Pi_{\Delta(\mathcal{A}^n)}$ is the Euclidean projection onto $\Delta(\mathcal{A}^n)$. We use these as useful comparison dynamics, as they are not GBAs under Definition 4. Euclidean FTRL can assign zero probability to an action whose gap remains of the same order as the tolerance, while zero-gap actions can remain forever unexplored under FTL and PGD, contradicting the small-gap exploration property. Here, FTL is an MBA but not a GBA and PGD is neither.

4. Large Gap Pruning: Best-Response Coarse Correlated Equilibria

We begin with the basic restriction imposed by large-gap pruning. If an action has positive static gap against the limiting opponent distribution, then its dynamic gap grows linearly. The GBA upper bound then makes its probability vanish, so it cannot have positive limiting frequency.

THEOREM 1 (Large Gap Pruning: Support Optimality). *Suppose $\bar{\mathbf{p}}_t \rightarrow \mathbf{p}$ and every player runs a GBA⁴. Then every action in a limiting marginal support is weakly optimal against the limiting opponent joint distribution: for every player n and action $a \in \mathcal{A}^n$,*

$$p^n(a) > 0 \implies \delta^n(a; \mathbf{p}) = 0.$$

Proof. We proceed by contradiction. Assume that action a has positive support $p^n(a) > 0$ but also positive static gap $\delta^n(a; \mathbf{p}) > 0$. Choose $\eta > 0$ such that

$$\max_{b \in \mathcal{A}^n} u^n(b, \mathbf{p}^{-n}) - u^n(a, \mathbf{p}^{-n}) > 2\eta.$$

⁴ Note that the proof of Theorem 1 only requires large-gap pruning Equation (Large Gap Pruning), and thus, also holds for MBAs. In particular, once $\Delta_t^n(a) > t\beta_t$, the MBA condition in Definition 5 gives $p_t^n(a) \leq \beta_t \rightarrow 0$.

Proposition 12 implies that $\Delta_t^n(a) > \eta t$ for all sufficiently large t . Since $\eta t \gg t^{\alpha_n}$, our large gap pruning property and Equation (7) give $\gamma_t^n(\Delta_t^n(a)) \rightarrow 0$ and hence $p_t^n(a) \rightarrow 0$. Therefore

$$\bar{p}_t^n(a) = \frac{1}{t} \sum_{\tau < t} p_\tau^n(a) \rightarrow 0,$$

contradicting $\bar{p}_t^n(a) \rightarrow p^n(a) > 0$. Thus, every supported action must have 0 static gap. $\square$

The support-optimality theorem (Theorem 1) identifies which actions can appear in each player's limiting support. To fully appreciate this result, we must first formally define the CR support restriction [9].

PROPOSITION 2 (Correlated-rationalizable and support-optimal sets). *Define the CR action set $\tilde{\mathcal{A}} \doteq (\tilde{\mathcal{A}}^1, \dots, \tilde{\mathcal{A}}^N)$ as the largest action set⁵ such that every $a \in \tilde{\mathcal{A}}^n$ is a best response to some $\mathbf{p}^{-n} \in \Delta(\tilde{\mathcal{A}}^{-n})$:*

$$\forall n, a \in \tilde{\mathcal{A}}^n, \quad \exists \mathbf{p}^{-n} \in \Delta(\tilde{\mathcal{A}}^{-n}) \quad \text{s.t.} \quad a \in \arg \max_{b \in \mathcal{A}^n} u^n(b, \mathbf{p}^{-n}).$$

Defining the support-optimal set $\mathcal{P}^* \doteq \{\mathbf{p} \in \Delta(\mathcal{A}) : p^n(a) > 0 \Rightarrow \delta^n(a; \mathbf{p}) = 0 \text{ for every } n, a\}$ —i.e., the set of distributions for which every supported action in $\mathbf{p}^n$ is optimal against $\mathbf{p}^{-n}$ —we have that $\mathcal{P}^* \subseteq \Delta(\tilde{\mathcal{A}})$.

Proof. Take any $\mathbf{p} \in \mathcal{P}^*$. The tuple of marginal supports $(\text{Supp}(\mathbf{p}^1), \dots, \text{Supp}(\mathbf{p}^N))$ itself satisfies the defining property of a CR action set: for every player n and every $a \in \text{Supp}(\mathbf{p}^n)$, support optimality gives $\delta^n(a; \mathbf{p}) = 0$, so a is a best response to $\mathbf{p}^{-n}$, which is supported on $\prod_{i \neq n} \text{Supp}(\mathbf{p}^i)$. Because $\tilde{\mathcal{A}}$ is the largest action set with this property, $\text{Supp}(\mathbf{p}^n) \subseteq \tilde{\mathcal{A}}^n$ for every player n . Hence $\mathbf{p} \in \Delta(\tilde{\mathcal{A}})$. $\square$

Now, we combine the support-optimal set $\mathcal{P}^*$ with no regret to obtain our newly proposed solution concept of BR-CCE—i.e., CCEs wherein each supported action is a best response to the opponents' joint distribution.

DEFINITION 6 (SUPPORT-REFINED CCE SETS). A joint distribution $\mathbf{p}$ is a CCE, written $\mathbf{p} \in \text{CCE}$, if $u^n(\mathbf{p}) \geq u^n(b, \mathbf{p}^{-n})$ for every player n and action $b \in \mathcal{A}^n$. It is a correlated equilibrium (CE) if, for every player n and actions $a, b \in \mathcal{A}^n$ we have $\sum_{\mathbf{a}^{-n}} p(a, \mathbf{a}^{-n}) (u^n(a, \mathbf{a}^{-n}) - u^n(b, \mathbf{a}^{-n})) \geq 0$. Famously, the empirical limits of dynamics with no external regret and no swap regret are CCEs

⁵ Since the game is finite, $\tilde{\mathcal{A}}$ is unique and can be found by repeatedly deleting actions that are not best responses to any distribution over the remaining opponent actions, as in previous pruning analyses of MBAs and CR [9, 11, 55, 96].

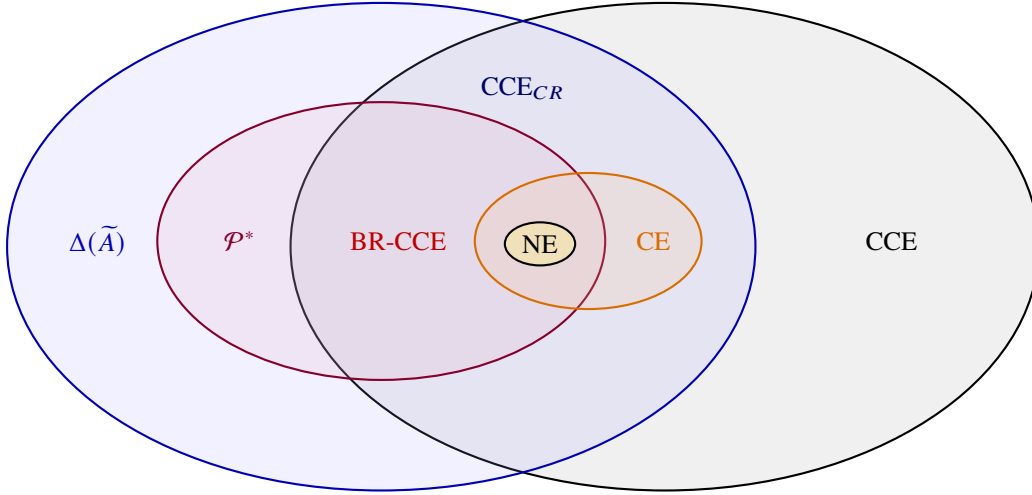


Figure 2 Relationship among CCE and the support refinements. CCE_{CR} denotes CCE supported on CR actions, while BR-CCE further requires each supported action to be optimal against the limiting opponent distribution.

and CEs, respectively [17]. Define CCE_{CR} and BR-CCE as the CR-restricted and support-optimal CCE sets, respectively:

$$CCE_{CR} \doteq CCE \cap \Delta(\tilde{\mathcal{A}}), \quad \text{BR-CCE} \doteq CCE \cap \mathcal{P}^*.$$

From Proposition 2, we are guaranteed that $NE \subseteq \text{BR-CCE} \subseteq CCE_{CR} \subseteq CCE$. In particular, every NE is a CCE and every action used by a NE is a best response to the opponents' marginal distribution, so $NE \subseteq \text{BR-CCE}$. The inclusion $\text{BR-CCE} \subseteq CCE_{CR}$ follows from Proposition 2, and $CCE_{CR} \subseteq CCE$ is immediate from Definition 6. Moreover, $CE \subseteq CCE_{CR}$ as each recommended action is optimal against the opponents' conditional distribution. We display the relationship between these sets in Figure 2.

PROPOSITION 3 (No-regret GBA limits are BR-CCE). *Suppose $\bar{\mathbf{p}}_t \rightarrow \mathbf{p}$. If every player uses an external-no-regret GBA, then $\mathbf{p} \in \text{BR-CCE}$.*

Proof. No-regret in Definition 5 implies $\mathbf{p} \in CCE$. Indeed, for any player n and fixed action b ,

$$u^n(b, \bar{\mathbf{p}}_t^{-n}) - u^n(\bar{\mathbf{p}}_t) = \frac{1}{t} \left(U_t^n(b) - \sum_{\tau < t} u^n(\mathbf{p}_\tau) \right) + O(1/t) \leq o(1).$$

Taking the limit along $\bar{\mathbf{p}}_t \rightarrow \mathbf{p}$ yields the CCE inequality in Definition 6. Theorem 1 gives the support-optimality condition, so $\mathbf{p} \in \text{BR-CCE}$ by Definition 6. $\square$

4.1. Strict Welfare Separation in Pay-as-Bid Auctions

Proposition 3 shows that every limit of external-no-regret GBA dynamics is a BR-CCE. This raises a substantive question: does replacing correlated-rationalizable support by support optimality only refine the set of possible limits, or does it also yield a strictly stronger welfare prediction? We answer this question by comparing the worst-case welfare of CCE, CCE_{CR} , BR-CCE, MNE, $\mathcal{P}^*$, and distributions supported on correlated-rationalizable actions. First, comparing $\Delta(\tilde{\mathcal{A}})$ with $\mathcal{P}^*$ isolates the contribution of support optimality alone, without imposing the CCE inequalities. Second, comparing CCE_{CR} with BR-CCE asks what the same support refinement contributes after no regret dynamics have already restricted the limit to CCE.

To make this comparison concrete but still tractable, we study the special case of the PAB auction with $N = 2$ bidders and $M = 2$ items (see Section 6.1 for the general N and M version). Each bidder $n \in \{1, 2\}$ has weakly decreasing positive integer marginal values $\mathbf{v}^n = (v_1^n, v_2^n)$, and we define the action set

$$\mathcal{A}^n(\mathbf{v}^n) \doteq \{\mathbf{b}^n \in \mathbb{Z}_{\geq 0}^2 : b_1^n \geq b_2^n, \quad b_m^n < v_m^n \text{ for } m \in \{1, 2\}\}. \quad (8)$$

The two largest unit bids win and, without loss of generality, bidder 1 has priority in every tie. Each bidder pays the sum of its winning bids. For a valuation profile $\mathbf{v}$, let $\text{WEL}(\mathbf{b}; \mathbf{v})$ denote the sum of the marginal values associated with the two winning bids, and let $\text{OPT}(\mathbf{v})$ denote the sum of the two largest values in $\{v_1^1, v_2^1, v_1^2, v_2^2\}$. For a distribution $\mathbf{p}$, write

$$\text{WEL}(\mathbf{p}; \mathbf{v}) \doteq \mathbb{E}_{\mathbf{b} \sim \mathbf{p}} [\text{WEL}(\mathbf{b}; \mathbf{v})]. \quad (9)$$

For any solution class $\mathcal{E} \in \{\text{CCE}, \text{CCE}_{CR}, \text{BR-CCE}, \text{MNE}, \mathcal{P}^*, \Delta(\tilde{\mathcal{A}})\}$, define its welfare price of anarchy by

$$\text{PoA}_{\mathcal{E}}(2, 2) \doteq \inf_{\mathbf{v}} \inf_{\mathbf{p} \in \mathcal{E}(\mathbf{v})} \frac{\text{WEL}(\mathbf{p}; \mathbf{v})}{\text{OPT}(\mathbf{v})}, \quad (10)$$

where for convenience, we will drop the common argument of $(2, 2)$ in $\text{PoA}_{\mathcal{E}}(2, 2)$ in this section. Larger PoA values correspond to stronger worst-case welfare guarantees. We give PoA bounds for each of these solution classes and show that the BR-CCE restriction strictly improves upon the CR restriction, with or without CCE constraints.

THEOREM 2 (PAB welfare hierarchy). *We have*

$$\text{PoA}_{\Delta(\tilde{\mathcal{A}})} = 0 < 11/14 = \text{PoA}_{\mathcal{P}^*}, \quad \text{PoA}_{\text{CCE}} \leq \text{PoA}_{\text{CCE}_{CR}} < \text{PoA}_{\text{BR-CCE}} = \text{PoA}_{\text{MNE}} = 5/6.$$

The left side of Theorem 2 isolates the contribution of support optimality. Restricting play only to CR actions gives no positive welfare guarantee, whereas support-optimality raises the tight guarantee to $11/14$. The right side shows what changes after imposing the CCE inequalities: a CCE supported on CR actions can still have welfare strictly below $5/6$. By contrast, BR-CCE achieves a tight $5/6$ guarantee. We prove this in Appendix B.3, though we note here that tight guarantees for general numbers of bidders N or units M , as well as the exact PoA of CCE and CCE_{CR} , remain open problems.

Implications for learning dynamics. It is known that no external regret dynamics converge towards the CCE set [17] and MBA dynamics converge to the CR set [9]. However, we are *not* guaranteed that large gap pruning in GBAs alone can yield convergence to $\mathcal{P}^*$. Thus, the aforementioned stronger welfare guarantees of $\mathcal{P}^*$ and BR-CCE apply *only if* the empirical distribution converges to $\mathcal{P}^*$. As such, these welfare improvements are conditional upon empirical convergence: support-optimal convergent dynamics yield stronger welfare guarantees than non-convergent dynamics.

5. Small Gap Exploration: Substitute-Driven Instability

Large-gap pruning rules out limiting support on strictly suboptimal actions. Small-gap exploration addresses the opposite question: when can a substitute—which is optimal against the limiting opponent distribution but absent from the limiting support—remain absent also from the dynamics? To answer this question, we must understand which opponent action profiles incentivize an agent to remain at equilibrium over deviating to this substitute.

DEFINITION 7 (STAY PROFILES). For a distribution $\mathbf{p}$ and substitute $a \in \mathcal{S}^n(\mathbf{p})$, define

$$\text{STAY}^{-n}(a; \mathbf{p}) \doteq \{\mathbf{a}^{-n} : u^n(b, \mathbf{a}^{-n}) > u^n(a, \mathbf{a}^{-n}) \text{ for some } b \in \text{Supp}(\mathbf{p}^n)\}$$

to be the set of opponent profiles on which at least one action in player n 's limiting support $\mathbf{p}^n$ strictly outperforms the substitute a .

Now, we state the main theorem guaranteed by GBAs' small-gap exploration.

THEOREM 3 (Small Gap Exploration: Necessary Stay Frequency). *Let $\mathbf{p}$ be a (possibly correlated) distribution and $a \in \mathcal{S}^n(\mathbf{p})$ be a substitute for player n . If GBA dynamics converge to $\mathbf{p}$, then*

$$P_t^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \gg t^{\alpha_n}.$$

Theorem 3 requires the cumulative stay frequency to eventually exceed the tolerance. To see this, if it does *not*, then small-gap exploration gives $a \in \mathcal{S}^n(\mathbf{p})$ eventually positive probability, contradicting its absence from the limiting support. Importantly, neither Theorem 1 nor Theorem 3 requires $\mathbf{p}$ to be a Nash equilibrium. For any possibly correlated empirical GBA limit, every supported action is optimal against the limiting opponent distribution, and the stay frequency of every substitute exceeds the tolerance. We prove this formally in Appendix C.1.

As we show in the next section, the case when the set of stay profiles is empty guarantees that its frequency is trivially below tolerance. After this, we give the forgiving condition on the dynamics and the candidate distribution $\mathbf{p}$ under which the stay frequency is below tolerance. To illustrate the power of our method, we rule out no-trade limits in bilateral trade and the low-bid weak PNE in the first-price auction.

5.1. Empty STAY Case

If no opponent profile allows any action in the limiting marginal support to strictly outperform the substitute, the required stay frequency is zero.

COROLLARY 1 (Empty stay sets rule out convergence). *Let $\mathbf{p}$ be a distribution and let $a \in \mathcal{S}^n(\mathbf{p})$. If $\text{STAY}^{-n}(a; \mathbf{p}) = \emptyset$, then GBA dynamics cannot converge to $\mathbf{p}$.*

Proof. Empty sets have frequency 0, contradicting convergence to $\mathbf{p}$ as per Theorem 3. $\square$

Now, we apply the above result to show that GBA dynamics guarantee divergence away from no-trade outcomes in bilateral-trade games.

Example: Bilateral Trade. A buyer (player B) has value $v \in \mathbb{N}$ and a seller (player S) has cost $c < v$ for some $c \in \mathbb{N}$. The buyer bids $b \in \{0, \dots, v-1\}$ and the seller asks $p \in \{c+1, \dots, V\}$ for some $V \geq v$; trade occurs if $b \geq p$ at midpoint price $(b+p)/2$, and the buyer's and seller's utilities are given by $v - (b+p)/2$ and $(b+p)/2 - c$ respectively. No trade occurs if $b < p$, and both agents obtain 0 utility. In particular, there exist NE supported exclusively over bids $b \leq c$ and asks $p \geq v$. We show the stronger statement that no possibly correlated limit can be supported on these no-trade profiles.

PROPOSITION 4 (No-trade distributions cannot be limits). *In the bilateral trade game with $v > c+1$, GBA dynamics cannot converge to any possibly correlated distribution $\mathbf{p}$ satisfying*

$$\text{Supp}(\mathbf{p}^B) \subseteq \{0, \dots, c\} \quad \text{and} \quad \text{Supp}(\mathbf{p}^S) \subseteq \{v, \dots, V\}.$$

Proof. Fix any distribution $\mathbf{p}$ with the stated supports and consider the buyer bid $v - 1$. It is unsupported because every action in $\text{Supp}(\mathbf{p}^B)$ is at most $c < v - 1$. It is also a best response to $\mathbf{p}^S$: every supported seller ask is at least v , so every buyer action gives payoff zero. Thus $v - 1 \in \mathcal{S}^B(\mathbf{p})$ by Definition 2. Moreover, no action in $\text{Supp}(\mathbf{p}^B)$ strictly outperforms $v - 1$ against any feasible seller ask. If the ask is at least v , both bids give zero; if it lies in $\{c + 1, \dots, v - 1\}$, the bid $v - 1$ trades at positive surplus while every supported buyer bid loses. Hence $\text{STAY}^S(v - 1; \mathbf{p}) = \emptyset$ by Definition 7, and Corollary 1 rules out convergence to $\mathbf{p}$. The seller-side argument is symmetric: the ask $c + 1$ is an unsupported best response with empty stay set. $\square$

Proposition 4 guarantees that GBA dynamics cannot converge to a no-trade profile, or to any distribution over them. However, we show that no-regret or MBA dynamics do not give the same guarantee.

PROPOSITION 5 (MBAs and no-regret dynamics can converge to no trade). *There exist MBA dynamics and no-regret dynamics whose play converges to the endpoint no-trade equilibrium (c, v) in the bilateral-exchange game.*

Proof overview. A deterministic tie-breaking MBA can choose bid c for the buyer and ask v for the seller whenever all gaps are below the MBA threshold. Against (c, v) , every action of both players receives payoff zero, so all gaps remain zero and the rule stays at no trade. A no-regret construction uses PGD initialized at the same endpoint actions; because every action earns zero against the endpoint opponent action, the PGD update remains fixed and regret is zero along this path. Appendix D contains the full proof and the warm-up simulations comparing GBA dynamics with MBA and no-regret comparison dynamics.

5.2. Forgiving Dynamics

Now, we consider the case when the stay set is non-empty. Theorem 3 reduces the problem to bounding its cumulative frequency. Unlike the empty stay set case, we will require an additional—albeit quite natural—condition on the algorithms which we call *forgiveness*.

DEFINITION 8 (FORGIVING). The dynamics are *n-forgiving* if, for every $i \neq n$ and every $\kappa > 0$,

$$\sum_{\tau < t} \gamma_{\tau}^i(\kappa \tau) \lesssim t^{\alpha_n}.$$

The dynamics are said to be *forgiving* if they are *n-forgiving* for every player n .

Definition 8 says that each opponent plays its linear-gap actions within player n 's tolerance. For example, it holds for every α_n whenever $\gamma_t^i(\kappa t) = O(1/t)$, since linear-gap actions then have only logarithmic cumulative frequency. Table 2 lists algorithm families satisfying this condition, such as Hedge and Tsallis FTRL. We now show that forgiving dynamics keep the stay frequency within tolerance whenever every stay profile contains an opponent action with positive static gap.

THEOREM 4 (Non-Convergence of Forgiving Dynamics to Certain Limits). *Let $\mathbf{p}$ be a distribution and let $\mathbf{a} \in \mathcal{S}^n(\mathbf{p})$. Suppose that every profile in the stay set from Definition 7 contains an opponent action with positive static gap:*

$$\mathbf{a}^{-n} \in \text{STAY}^{-n}(\mathbf{a}; \mathbf{p}) \implies \delta^i(\mathbf{a}^i; \mathbf{p}) > 0 \text{ for some } i \neq n$$

Then n -forgiving GBA dynamics cannot converge to $\mathbf{p}$.

The above result is our main theoretical contribution of this section. In particular, both the forgiving algorithmic condition and the corresponding equilibrium conditions required for this non-convergence result can be verified independently of the exact choice of dynamics or initialization bias. The idea is simple. An opponent action with positive static gap has a dynamic gap that grows linearly (Proposition 12), so forgiveness keeps its cumulative frequency within player n 's tolerance. Every stay profile contains such an action, so the stay frequency stays within tolerance as well, which contradicts Theorem 3. We give the full proof in Appendix C.2.

5.3. Algorithm Families and the FPA

The forgiving theorem combines an algorithmic condition with an equilibrium condition on the candidate limit. Table 2 summarizes the algorithmic condition for Hedge, polynomial weights, Tsallis and log-barrier FTRL (regularizers in Table 4), and buffered polynomial weights. Hedge is always forgiving under the parameterization in Section 3; for the other families, forgiveness depends on the rate at which linear-gap actions are pruned. Proposition 9 in Appendix A.2 gives the formal statements and proofs.

Algorithm family	n -forgiving
Hedge / entropy FTRL	Always
Polynomial weights / Tsallis FTRL / log-barrier FTRL	If $\beta_i(1 - \alpha_i) \geq 1 - \alpha_n$ for every $i \neq n$
Buffered polynomial weights, $\epsilon_t^i = t^{-s_i}$	If $\min\{\beta_i(1 - \alpha_i), s_i\} \geq 1 - \alpha_n$ for every $i \neq n$
FTL / Euclidean FTRL / PGD	Not a GBA

Table 2 Algorithmic conditions for forgiveness. Section 3 defines the parameters: α_i is the GBA tolerance exponent, β_i is the polynomial-tail exponent, and s_i is the buffered exploration-floor exponent. Tsallis FTRL with regularizer parameter $q_{ts} \in (0, 1)$ has tail exponent $\beta_i = 1/(1 - q_{ts})$, and log-barrier FTRL has $\beta_i = 1$, so both are forgiving when all players share a tolerance exponent. We give the formal justification in Proposition 9.

We now use the FPA with 2 bidders tied with the largest valuation to make our forgiving equilibrium condition concrete.

Example: FPA. Consider the first-price auction with two bidders with common valuation $v \geq 2$. Bids are integers in $\{0, \dots, v-1\}$, the highest bidders split the item, and each winner pays its own bid. As shown by [22], the two-top-bidder subgame has a weak PNE at $b^1 = b^2 = v-2$ with substitute $v-1$, and a strict PNE at $b^1 = b^2 = v-1$.

COROLLARY 2 (FPA equilibrium selection). *In the FPA, forgiving GBA dynamics cannot converge to the weak PNE $b^1 = b^2 = v-2$. Consequently, if these dynamics converge to one of the two PNE described above, the limit must be the high-bid PNE $b^1 = b^2 = v-1$.*

Proof. At the weak PNE $b^1 = b^2 = v-2$, the unsupported best response $v-1$ is a substitute by Definition 2. For either bidder n , Definition 7 gives

$$\text{STAY}^{-n}(v-1; v-2) = \{0, 1, \dots, v-3\}.$$

Every bid in this stay set has positive static gap for the opponent, because it loses while $v-2$ obtains payoff one. Theorem 4 therefore rules out convergence under forgiving dynamics. Finally, [22] shows that the two PNE in this setting are the low-bid and high-bid profiles described above. Since the low-bid PNE has been ruled out, any limit that is one of these PNE must be the high-bid PNE. $\square$

PROPOSITION 6 (Weaker dynamics can converge to the weak FPA equilibrium). *In the two-bidder FPA, there exist MBA dynamics, no-regret dynamics, and non-forgiving GBA dynamics whose play converges to the weak PNE $b^1 = b^2 = v-2$.*

Proof overview. The construction follows the two-highest-value-bidder analysis of [22]. For MBA dynamics, FTL with a finite prior favoring $v-2$ keeps $v-2$ as the unique cumulative leader because $v-2$ and $v-1$ earn the same payoff against $v-2$. For no-regret dynamics, PGD initialized at the vertex $v-2$ remains fixed there by the simplex projection calculation. For a non-forgiving GBA, both bidders run Hedge with tolerance t^α plus a small amount of uniform exploration, $\epsilon_t = ct^{-s}$ with $0 < s < 1 - \alpha$ and a small constant $c > 0$, starting from a prior that favors $v-2$. Whenever the opponent explores a low bid, $v-2$ gains on the substitute $v-1$, and this lead grows faster than the tolerance t^α , so the substitute is pruned. Appendix D.2 gives the details, and Table 5 reports simulations of these and related dynamics.

Limitations of Small Gap Exploration: Appendix C.3 notes two limitations of the forgiving condition. First, not every GBA is forgiving. Second, not every weak NE satisfies the corresponding static condition. Thus Theorem 4 is a sufficient non-convergence test rather than a complete characterization of weak NE.

6. Multi-Unit Auction Applications and Numerical Experiments

Sections 4 and 5 give complementary restrictions on the limits of GBA dynamics. We begin by defining PAB and UPA in Section 6.1. Section 6.2 then compares the welfare and revenue of CCE, CCE_{CR} , BR-CCE, and Hedge. The experiments show that BR-CCE is consistently more informative than CCE and CCE_{CR} , especially for PAB. For UPA, however, Hedge approaches zero revenue even though the three static solution concepts permit positive-revenue outcomes, consistent with prior simulations [32]. This gap motivates the theoretical analysis in Section 6.3, where Theorem 5 uses the forgiving condition to rule out convergence to positive-revenue PNEs under forgiving GBA dynamics.

6.1. Auction formats

Pay-as-bid. In the PAB multi-unit auction, there are N bidders competing for M identical items. Each bidder n has a nonincreasing marginal value vector $\mathbf{v}^n = (v_1^n, \dots, v_M^n)$. Individual bids are ranked and the M highest unit bids win—with ties broken in favor of lower indexed bidders for concreteness—and each bidder’s payment is equal to the sum of its winning bids.

Uniform-price auctions. The UPA uses the same allocation rule but a common per-unit price equal to the highest rejected bid. Revenue is the common price times M . Each bidder may submit any nonincreasing, non-negative integer bid vector $\mathbf{b}^n \in \mathbb{Z}_{\geq 0}^M$ satisfying no overbidding—i.e., $b_m^n < v_m^n$.

Experimental protocol. Unless otherwise stated, simulations use expected-feedback updates with independent play. In each random instance, marginal values are integers in $\{1, \dots, V\}$. Welfare and revenue are normalized by the efficient welfare benchmark OPT, equal to the sum of the top M realized marginal values. The experiments below compare CCE, CCE_{CR} , exact BR-CCE, and Hedge dynamics; Appendix A.2 gives the formal update rules and discusses the broader algorithm families used elsewhere in the paper. Appendix E records the corresponding computational details.

6.2. BR-CCE experiments for multi-unit auctions

We next compare CCE, CCE_{CR} , exact BR-CCE, and Hedge dynamics for the two multi-unit payment rules. The static rows optimize three objectives separately: minimum welfare, maximum

revenue, and minimum revenue. We report $1 - \text{WEL}$ rather than welfare because most welfare values are close to one. The two revenue objectives answer different questions. The maximum-revenue objective tests whether BR-CCE already eliminates positive-revenue UPA outcomes; it does not. The minimum-revenue objective tests whether BR-CCE lifts the low-revenue part of the equilibrium set and whether the refined benchmark is closer to the learning dynamics, especially for PAB.

Table 3 uses $N = 2$, $M = 3$, $V = 8$, and 25 random instances. We use this configuration because exact BR-CCE optimization is computationally expensive. In particular, while the CCEs can be computed by linear programming and the CR action set can be computed by the iterative procedure described after Proposition 2, the BR-CCE were computed via a mixed-integer support formulation. In the table, BR-CCE is closest to the PAB Hedge dynamics, especially in minimum revenue and welfare loss. For UPA, BR-CCE improves over CCE and CCE_{CR} modestly, but it still permits both zero-revenue and positive-revenue outcomes. This suggests that large-gap pruning plus CCE does not by itself explain the low-revenue selection observed under UPA dynamics.

Table 3 Multi-unit auction experiments for $N = 2$, $M = 3$, $V = 8$ over 25 random instances. Entries are mean/worst normalized by OPT. For $1 - \text{WEL}$, worst means largest welfare loss; for $\text{REV}_{\max}$, worst means largest maximum revenue; and for $\text{REV}_{\min}$, worst means smallest minimum revenue. Static benchmark rows are shaded in gray. For both Hedge rows, $\text{REV}_{\max}$ worst entry is the largest realized revenue across instances, while its $\text{REV}_{\min}$ worst entry is the smallest realized revenue.

Rule	Object	$1 - \text{WEL}$	$\text{REV}_{\max}$	$\text{REV}_{\min}$
PAB	CCE	0.021 / 0.119	0.255 / 0.429	0.183 / 0.000
PAB	CCE_{CR}	0.020 / 0.103	0.254 / 0.429	0.183 / 0.000
PAB	BR-CCE	0.009 / 0.083	0.254 / 0.429	0.233 / 0.000
PAB	Hedge	0.007 / 0.072	0.241 / 0.426	0.241 / 0.000
UPA	CCE	0.022 / 0.122	0.123 / 0.429	0.000 / 0.000
UPA	CCE_{CR}	0.015 / 0.099	0.123 / 0.429	0.000 / 0.000
UPA	BR-CCE	0.008 / 0.071	0.117 / 0.429	0.000 / 0.000
UPA	Hedge	0.000 / 0.001	0.038 / 0.253	0.038 / 0.000

6.3. UPA

The UPA rows of Table 3 show that positive-revenue BR-CCE outcomes remain feasible. In fact, many of the positive-revenue BR-CCE outcomes found in those instances were PNEs. Thus large-gap pruning, even combined with CCE, cannot by itself rule out the positive-revenue UPA equilibria that remain feasible in the static analysis. This is where Theorem 4 becomes useful. In particular, UPA is our main application of small-gap exploration. Here, we identify a price-lowering substitute that has zero static gap and therefore is not removed by large-gap pruning; under forgiving GBA dynamics, its continued exploration destabilizes positive-revenue PNEs.

THEOREM 5 (Positive-revenue UPA PNE cannot be limits). *Consider UPA with integer bids and no overbidding such that:*

1. *Each bidder's valuations are generic: no bidder is indifferent between any two bids with different allocations—e.g., valuations are drawn from an absolutely continuous density.*
2. *Each bidder is non-spectating: bidder n guarantees allocation by bidding $\mathbf{b}^n = \lceil \mathbf{v}^n \rceil - \mathbf{1}$, its highest feasible bid (here $\mathbf{1}$ is the all-ones vector).*

Then, forgiving GBA dynamics cannot converge to any positive-revenue PNE.

Proof overview. Fix a positive-revenue PNE $\mathbf{b}$ with clearing price $\pi > 0$. First, every bidder n must receive a positive allocation, otherwise they can bid $\mathbf{b}^n = \lceil \mathbf{v}^n \rceil - \mathbf{1}$ which by our non-spectating assumption guarantees strictly positive surplus. Consequently, at least two bidders must have losing bids equal to π : if one bidder were the unique price setter, it could lower its losing π -bids, preserve its allocation, and strictly reduce its payment. Fix one such price-setting bidder n , and let $\mathbf{b}_*^n$ lower all of its losing π -bids to $\pi - 1$, leaving its winning bids unchanged. Lemma 2 shows that this does not change bidder n 's allocation. Since another bidder still has a losing bid equal to π , the clearing price also remains π when the opponents play $\mathbf{b}^{-n}$. Hence $u^n(\mathbf{b}_*^n, \mathbf{b}^{-n}) = u^n(\mathbf{b}^n, \mathbf{b}^{-n})$, so $\mathbf{b}_*^n$ is a substitute. It remains to verify the condition of Theorem 4. Consider an opponent profile $\mathbf{a}^{-n}$ in which every opponent uses a zero-static-gap action. By our genericity assumption, each such action gives that opponent the same allocation as at the PNE. Let m_i be bidder i 's PNE allocation. For any bid lowered in $\mathbf{b}_*^n$, the opponents' zero-gap actions contribute $\sum_{i \neq n} m_i$ bids ranked above it, while bidder n 's unchanged winning bids contribute another m_n . Since $\sum_i m_i = M$, every lowered bid remains losing even when all zero-gap opponent actions are played jointly. Lowering these bids therefore preserves bidder n 's allocation and can only weakly reduce the clearing price: $u^n(\mathbf{b}_*^n, \mathbf{a}^{-n}) \geq u^n(\mathbf{b}^n, \mathbf{a}^{-n})$. Thus no profile consisting entirely of zero-static-gap opponent actions belongs to $\text{STAY}^{-n}(\mathbf{b}_*^n; \mathbf{b})$. Every stay profile must contain an opponent action with positive static gap, so Theorem 4 rules out convergence to $\mathbf{b}$. Appendix E.3 gives the full proof of Theorem 5.⁶

PROPOSITION 7 (Weaker dynamics can remain at a positive-revenue UPA PNE). *There exists a UPA instance with a positive-revenue PNE at which MBA and no-regret learning dynamics can remain from suitable initializations.*

⁶ Our integer-valued experiments are not generic, but subtracting a tiny amount from every value makes them generic without noticeably changing the simulations; see Appendix E.3.

Proof overview. We give a concrete example in Appendix E.4 with a two-item, two-bidder UPA instance with a positive-revenue PNE. We verify the updates directly, showing convergence. For FTL, a finite initialization makes each PNE bid the unique cumulative leader, and the PNE property ensures that no alternative ever catches up once play remains at the PNE. For PGD, the initialization at the PNE bid never changes as it is a best response.

We note one important limitation of our UPA theorem, which is that our result does *not* rule out the possibility of positive-revenue MNEs or CCEs. Nonetheless, it offers a learning-based explanation for unusually low clearing prices in real-world UPA implementations in emissions-permit markets, where prices can be low relative to estimates of marginal climate damages or the social cost of carbon [13, 14, 53, 54, 93]. Since allowance-auction revenues fund energy-efficiency, renewable-energy, greenhouse-gas-reduction, bill-assistance, public-health, and environmental programs [50, 81, 86], the revenue-selection issue is economically meaningful. Figure 1 exhibits the same low-revenue pattern that motivates Theorem 5.

7. Extensions and Additional Considerations

Finite priors and time shifts. Our definitions allow finite initialization terms in cumulative utilities. Appendix A.3 shows formally that finite priors, warm starts, and deleting a finite prefix of play perturb dynamic gaps only by a bounded amount. Since our large-gap and small-gap arguments compare gaps to powers of time or to linear growth, these finite changes do not affect the asymptotic conclusions.

Realized full-information feedback. The analysis in Sections 2–6 uses expected full-information feedback for clarity. Proposition 11 gives high-probability static-gap bounds under realized full-information feedback, using the same concentration-inequality approach used in realized-feedback analyses of MBA learning [9, 22]. When every tolerance exponent satisfies $\alpha_n > 1/2$, Corollary 3 guarantees that the $O(\sqrt{t \log t})$ error is eventually $o(t^{\alpha_n})$ almost surely. It therefore extends Theorems 1, 3, and 4 to realized full-information feedback.

Computational hardness of BR-CCE. Optimizing over CCE_{CR} is a linear program once the CR action set is found, while optimizing over BR-CCE requires a nonconvex support condition. Appendix B gives the two-player support-collapse result and explains why BR-CCE optimization contains NE support selection as a special case. Since computing a Nash equilibrium in a general finite game is complete for Polynomial Parity Arguments on Directed graphs (PPAD) [21], this explains why our BR-CCE computations use mixed-integer support formulations.

8. Conclusion

Market games often contain broad CCE sets and weak NE sustained by substitutes. MBAs are too permissive to predict whether those actions remain in the limiting support, while no-regret/FTRL stability theory is often more structured than necessary. GBAs provide a middle ground through two complementary properties. Large-gap pruning refines empirical no-regret limits from CCE to BR-CCE by requiring every marginally used action to be weakly optimal against the empirical opponent distribution. Small-gap exploration prevents an unsupported substitute from disappearing unless the profiles that favor the equilibrium strategy occur frequently enough; forgiveness gives a simple sufficient condition under which they do not.

The main conceptual message is that weak NE differ in how they respond to substitute exploration. In the FPA, bilateral trade, and UPA, the profiles that favor the equilibrium action require opponents to use positive-static-gap actions. Forgiving dynamics prune those actions too quickly to eliminate the substitute. The example in Proposition 14 marks the boundary of this argument: not every weak NE satisfies the forgiving static condition, so some weak NE remain outside our non-convergence test.

The two multi-unit auction applications give the main market-design takeaway. In PAB, the same support-optimality theorem (Theorem 1) that underlies BR-CCE is highly predictive: worst BR-CCE outcomes and Hedge dynamics both deliver high welfare and substantially higher revenue than the worst CCE benchmark. In UPA, by contrast, losing bids can set a common price, so exploration of price-lowering substitutes under forgiving GBA dynamics destabilizes positive-revenue PNEs and the simulations move toward low revenue. This contrast suggests a practical lesson for market designers: the payment rule determines whether substitute exploration erodes prices, as in UPA, or instead leaves the GBA/no-regret refinement close to high-welfare and high-revenue outcomes, as in PAB.

References

- [1] Ahunbay MS, Bichler M (2025) Semicoarse correlated equilibria and LP-based guarantees for gradient dynamics in normal-form games. URL <https://arxiv.org/abs/2502.20466>.
- [2] Ahunbay MŞ, Zheng W, Lin T (2026) Gradient dynamics in first-price auctions: Iterative strategy elimination via cubic potentials. URL <https://arxiv.org/abs/2606.05108>.
- [3] Anagnostides I, Panageas I, Farina G, Sandholm T (2022) On last-iterate convergence beyond zero-sum games. Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, eds., *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, 536–581 (PMLR), URL <https://proceedings.mlr.press/v162/anagnostides22a.html>.
- [4] Ausubel LM, Cramton P (2006) Dynamic auctions in procurement. Dimitri N, Piga G, Spagnolo G, eds., *Handbook of Procurement*, 220–246 (Cambridge University Press), URL <http://dx.doi.org/10.1017/CBO9780511492556.010>.
- [5] Balseiro S, Golrezaei N, Mahdian M, Mirrokni V, Schneider J (2022) Contextual bandits with cross-learning. *Mathematics of Operations Research* 48(3):1607–1629, URL <http://dx.doi.org/10.1287/moor.2022.1313>.
- [6] Balseiro SR, Gur Y (2019) Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science* 65(9):3952–3968, URL <http://dx.doi.org/10.1287/mnsc.2018.3174>.

- [7] Bergemann D, Brooks B, Morris S (2017) First-price auctions with general information structures: Implications for bidding and revenue. *Econometrica* 85(1):107–143, ISSN 00129682, 14680262, URL <http://www.jstor.org/stable/44155418>.
- [8] Bernasconi M, Castiglioni M, Celli A, Fusco F (2024) No-regret learning in bilateral trade via global budget balance. *STOC 2024: Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, URL <http://dx.doi.org/10.1145/3618260.3649653>.
- [9] Bichler M, Durmann J, Oberlechner M (2024) Online optimization algorithms in repeated price competition: Equilibrium learning and algorithmic collusion. URL <https://arxiv.org/abs/2412.15707>.
- [10] Binmore K, Swierzbinski J (2000) Treasury auctions: Uniform or discriminatory? *Review of Economic Design* 5:387–410.
- [11] Braverman M, Mao J, Schneider J, Weinberg SM (2017) Selling to a no-regret buyer. *CoRR* abs/1711.09176, URL <http://arxiv.org/abs/1711.09176>.
- [12] Brânzei S, Derakhshan M, Golrezaei N, Han Y (2023) Online learning in multi-unit auctions. URL <https://arxiv.org/abs/2305.17402v1>.
- [13] Burtraw D, Holt C, Palmer K, Paul A, Shobe W (2018) Quantities with prices. Technical Report Working Paper 18-08, Resources for the Future, URL <https://media.rff.org/documents/RFF20WP2018-08.pdf>.
- [14] Bushnell JB (2017) (overly) great expectations: Carbon pricing and revenue uncertainty in california. *National Tax Journal* 70(4):837–854, URL <http://dx.doi.org/10.17310/ntj.2017.4.07>.
- [15] California Legislative Analyst's Office (2025) Assessing California's climate policies—cap-and-trade reauthorization. Technical report, California Legislative Analyst's Office, URL <https://lao.ca.gov/Publications/Report/5042>.
- [16] Cesa-Bianchi N, Cesari T, Colomboni R, Fusco F, Leonardi S (2024) Bilateral trade: A regret minimization perspective. *Mathematics of Operations Research* 49(1):171–203, URL <http://dx.doi.org/10.1287/moor.2023.1351>.
- [17] Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (USA: Cambridge University Press), ISBN 0521841089.
- [18] Chen X, Peng B (2020) Hedging in games: Faster convergence of external and swap regrets. URL <https://arxiv.org/abs/2006.04953>.
- [19] Chen X, Peng B (2023) Complexity of equilibria in first-price auctions under general tie-breaking rules. URL <https://arxiv.org/abs/2303.16388>.
- [20] Daskalakis C, Fishelson M, Golowich N (2023) Near-optimal no-regret learning in general games. URL <https://arxiv.org/abs/2108.06924>.
- [21] Daskalakis C, Goldberg PW, Papadimitriou CH (2009) The complexity of computing a nash equilibrium. *Commun. ACM* 52(2):89–97, ISSN 0001-0782, URL <http://dx.doi.org/10.1145/1461928.1461951>.
- [22] Deng X, Hu X, Lin T, Zheng W (2022) Nash convergence of mean-based learning algorithms in first price auctions. *Proceedings of the ACM Web Conference 2022*, WWW '22 (ACM), URL <http://dx.doi.org/10.1145/3485447.3512059>.
- [23] Edelman B, Ostrovsky M, Schwarz M (2007) Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *American Economic Review* 97(1):242–259, URL <http://dx.doi.org/10.1257/aer.97.1.242>.
- [24] Fabra N, von der Fehr NH, Harbord D (2006) Designing electricity auctions. *The RAND Journal of Economics* 37(1):23–46, ISSN 07416261, URL <http://www.jstor.org/stable/25046225>.
- [25] Federico G, Rahman D (2003) Bidding in an electricity pay-as-bid auction. *Journal of Regulatory Economics* 24(2):175–211, URL <http://dx.doi.org/10.1023/A:1024738128115>.
- [26] Feldman M, Lucier B, Nisan N (2017) Correlated and coarse equilibria of single-item auctions. URL <https://arxiv.org/abs/1601.07702v3>.
- [27] Feng Y, Golrezaei N, Li Q (2026) Nash equilibria in uniform price auctions: Theory, computation, and market insights. *Abstracts of the 2026 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, 111–113 (Association for Computing Machinery), URL <http://dx.doi.org/10.1145/3801489.3806860>.
- [28] Feng Z, Guruganesh G, Liaw C, Mehta A, Sethi A (2020) Convergence analysis of no-regret bidding algorithms in repeated auctions. *CoRR* abs/2009.06136, URL <https://arxiv.org/abs/2009.06136>.
- [29] Filos-Ratsikas A, Giannakopoulos Y, Hollender A, Kokkalis C (2024) On the computation of equilibria in discrete first-price auctions. URL <https://arxiv.org/abs/2402.12068>.
- [30] Filos-Ratsikas A, Giannakopoulos Y, Hollender A, Lazos P, Poças D (2023) On the complexity of equilibrium computation in first-price auctions. *SIAM Journal on Computing* 52(1):80–131, ISSN 1095-7111, URL <http://dx.doi.org/10.1137/21m1435823>.
- [31] Foster DJ, Li Z, Lykouris T, Sridharan K, Tardos E (2016) Learning in games: Robustness of fast convergence. Lee D, Sugiyama M, Luxburg U, Guyon I, Garnett R, eds., *Advances in Neural Information Processing Systems*, volume 29 (Curran Associates, Inc.), URL https://proceedings.neurips.cc/paper_files/paper/2016/file/b3f61131b6ceeb2b14835fa648a48ff-Paper.pdf.
- [32] Galgana R, Golrezaei N (2025) Learning in repeated multiunit pay-as-bid auctions. *Manufacturing & Service Operations Management* 27(1):200–229, URL <http://dx.doi.org/10.1287/msom.2023.0403>.
- [33] Gallego G, Huh WT, Kang W, Phillips R (2006) Price competition with the attraction demand model: Existence of unique equilibrium and its stability. *Manufacturing & Service Operations Management* 8(4):359–375, URL <http://dx.doi.org/10.1287/msom.1060.0115>.
- [34] Gao X, Liu X, Chen R, Jiang Y (2025) A dynamic auction-based algorithm for computing constrained walrasian equilibrium. *2025 6th International Conference on Computer Science and Management Technology*, URL <http://dx.doi.org/10.1145/3795154.3795299>.
- [35] Garbade KD, Ingber JF (2005) The treasury auction process: Objectives, structure, and recent adaptations. *Current Issues in Economics and Finance* 11(2), URL https://www.newyorkfed.org/research/current_issues/ci11-2.html.
- [36] Giannou A, Vlatakis-Gkaragkounis EV, Mertikopoulos P (2021) Survival of the strictest: Stable and unstable equilibria under regularized learning with partial information. Belkin M, Kpotufe S, eds., *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, 2147–2148 (PMLR), URL <https://proceedings.mlr.press/v134/giannou21a.html>.
- [37] Goke S, Weintraub GY, Mastromonaco R, Seljan S (2022) Bidders' responses to auction format change in internet display advertising auctions. URL <https://arxiv.org/abs/2110.13814>.
- [38] Golowich N, Pattathil S, Daskalakis C (2020) Tight last-iterate convergence rates for no-regret learning in multi-player games. *CoRR* abs/2010.13724, URL <https://arxiv.org/abs/2010.13724>.
- [39] Golrezaei N, Jaillet P, Liang JCN, Mirrokni V (2023) Pricing against a budget and roi constrained buyer. *International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, 9282–9307 (PMLR), URL <https://proceedings.mlr.press/v206/golrezaei23a.html>.

- mlr.press/v206/golrezaei23a.html.
- [40] Golrezaei N, Javanmard A, Mirrokni V (2021) Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Operations Research* 69(1):297–314.
- [41] Golrezaei N, Sahoo S (2025) Learning safe strategies for value maximizing buyers in uniform price auctions. *International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, 19841–19877 (PMLR), URL <https://proceedings.mlr.press/v267/golrezaei25a.html>.
- [42] Golrezaei N, Sahoo S (2026) Learning to bid in discriminatory auctions with budget constraints. *The 29th International Conference on Artificial Intelligence and Statistics*, URL <https://openreview.net/forum?id=wtulrrJRT6>.
- [43] Goulder LH, Schein AR (2013) Carbon taxes versus cap and trade: A critical review. *Climate Change Economics* 4(3):1350010, URL <http://dx.doi.org/10.1142/S2010007813500103>.
- [44] Government of Québec (2026) The carbon market: Auctions. <https://www.environnement.gouv.qc.ca/changements/carbone/Ventes-encheres-en.htm>, accessed October 2026.
- [45] Government of Québec (2026) Québec and california current auction price history since november 2014. Technical report, Government of Québec, URL <https://www.environnement.gouv.qc.ca/changements/carbone/ventes-encheres/historique-prix-encheres-WCI-en.pdf>, historical auction-price series; accessed October 2026.
- [46] Guo MA, Ying D, Lavaei J, Shen ZJM (2023) No-regret learning in dynamic competition with reference effects under logit demand. *Advances in Neural Information Processing Systems*, URL https://papers.nips.cc/paper_files/paper/2023/file/21c9fd36a6d94a491bf330a0ba0e5f6e-Paper-Conference.pdf.
- [47] Han Y, Zhou Z, Flores A, Ordentlich E, Weissman T (2020) Learning to bid optimally and efficiently in adversarial first-price auctions. URL <http://dx.doi.org/10.48550/ARXIV.2007.04568>.
- [48] Han Y, Zhou Z, Weissman T (2020) Optimal no-regret learning in repeated first-price auctions. URL <http://dx.doi.org/10.48550/ARXIV.2003.09795>.
- [49] Han Z, Fang D, Yang P (2026) Managing coupled electricity and carbon markets: A multiagent equilibrium perspective. *IEEE Transactions on Engineering Management* 73:2597–2610, URL <http://dx.doi.org/10.1109/TEM.2026.3653209>.
- [50] International Carbon Action Partnership (2026) USA – california cap-and-invest program. <https://icapcarbonaction.com/en/ets/usa-california-cap-and-invest-program>, accessed June 5, 2026.
- [51] Jin Y, Lu P (2023) First price auction is $1 - 1/e^2$ efficient. *Journal of the ACM* 70(5):36:1–36:86, URL <http://dx.doi.org/10.1145/3617902>.
- [52] Klemperer P (2002) How (not) to run auctions: The european 3g telecom auctions. *European Economic Review* 46(4–5):829–845.
- [53] Koch N, Fuss S, Grosjean G, Edenhofer O (2014) Causes of the EU ETS price drop: Recession, CDM, renewable policies or a bit of everything?—new evidence. *Energy Policy* 73:676–685, URL <http://dx.doi.org/10.1016/j.enpol.2014.06.024>.
- [54] Kolstad CD (2020) Subnational carbon pricing in the us. *ifo DICE Report* 18(1):20–22.
- [55] Kolumbus Y, Nisan N (2022) Auctions between regret-minimizing agents. *Proceedings of the ACM Web Conference 2022*, 100–111, WWW ’22 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450390965, URL <http://dx.doi.org/10.1145/3485447.3512055>.
- [56] Kong XTR, He Z, Yu K, Yan P (2025) Optimal lot-sizing and service level weighting in sequential multi-attribute global transportation service procurement. *European Journal of Operational Research* 327(3):1052–1072, URL <http://dx.doi.org/10.1016/j.ejor.2025.05.037>.
- [57] Lebrun B (1996) Existence of an equilibrium in first price auctions. *Economic Theory* 7(3):421–443, ISSN 09382259, 14320479, URL <http://www.jstor.org/stable/25054935>.
- [58] Lebrun B (1999) First price auctions in the asymmetric n bidder case. *International Economic Review* 40(1):125–142, ISSN 00206598, 14682354, URL <http://www.jstor.org/stable/2648842>.
- [59] Lei Q, Nagarajan SG, Panageas I, Wang X (2021) Last iterate convergence in no-regret learning: constrained min-max optimization for convex-concave landscapes. Banerjee A, Fukumizu K, eds., *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, 1441–1449 (PMLR), URL <https://proceedings.mlr.press/v130/lei21a.html>.
- [60] Li M, Liu J, Wu XJ (2026) Machine behavior in multi-unit auctions. URL <https://ssrn.com/abstract=6725519>, available at SSRN.
- [61] Lin T, Chen Y (2024) Generalized Principal-Agent Problem with a Learning Agent. *arXiv e-prints* arXiv:2402.09721, URL <http://dx.doi.org/10.48550/arXiv.2402.09721>.
- [62] Liu J, Wu XJ, Li M (2026) AI behavior in repeated multi-unit auctions. URL <http://dx.doi.org/10.2139/ssrn.5332110>, uNSW Business School Research Paper Forthcoming.
- [63] Lopomo G, Marx LM, McAdams D, Murray B (2011) Carbon allowance auction design: An assessment of options for the united states. *Review of Environmental Economics and Policy* 5(1):25–43, URL <http://dx.doi.org/10.1093/reep/req024>.
- [64] Lunghi A, Castiglioni M, Marchesi A (2026) Better regret rates in bilateral trade via sublinear budget violation. *Proceedings of the 2026 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 6494–6536, URL <http://dx.doi.org/10.1137/1.9781611978971.233>.
- [65] Makhdoumi A, Pekeč S, Wang X (2026) Clearing multi-unit markets with simple pricing. URL <http://dx.doi.org/10.2139/ssrn.6228058>, available at SSRN.
- [66] Markakis E, Telelis O (2015) Uniform price auctions: Equilibria and efficiency. *Theor. Comp. Sys.* 57(3):549–575, ISSN 1432-4350, URL <http://dx.doi.org/10.1007/s00224-014-9537-9>.
- [67] Maskin E, Riley J (2000) Equilibrium in sealed high bid auctions. *The Review of Economic Studies* 67(3):439–454, ISSN 00346527, 1467937X, URL <http://www.jstor.org/stable/2566961>.
- [68] Milgrom P (2000) Putting auction theory to work: The simultaneous ascending auction. *Journal of Political Economy* 108(2):245–272, URL <http://dx.doi.org/10.1086/262118>.
- [69] Mohri M, Muñoz Medina A (2013) Learning theory and algorithms for revenue optimization in second-price auctions with reserve. URL <http://dx.doi.org/10.48550/ARXIV.1310.5665>.
- [70] Mohri M, Muñoz Medina A (2016) Learning algorithms for second-price auctions with reserve. *Journal of Machine Learning Research* 17(74):1–25, URL <https://www.jmlr.org/papers/v17/14-499.html>.

- [71] Nadav U, Piliouras G (2010) No regret learning in oligopolies: Cournot vs. bertrand. *Algorithmic Game Theory*, volume 6386 of *Lecture Notes in Computer Science*, 300–311 (Springer), URL http://dx.doi.org/10.1007/978-3-642-16170-4_26.
- [72] Nazir I, Mushtaq N, Amin W (2025) Smart grid systems: Addressing privacy threats, security vulnerabilities, and demand–supply balance (a review). *Energies* 18(19):5076, URL <http://dx.doi.org/10.3390/en18195076>.
- [73] Nedelec T, El Karoui N, Perchet V (2019) Learning to bid in revenue-maximizing auctions. *CoRR* abs/1902.10427, URL <http://arxiv.org/abs/1902.10427>.
- [74] Nisan N, Noti G (2017) An experimental evaluation of regret-based econometrics. *Proceedings of the 26th International Conference on World Wide Web*, 73–81, URL <http://dx.doi.org/10.1145/3038912.3052621>.
- [75] O'Connor T, Reyna E (2013) AB 32 cap-and-trade auction frequently asked questions. Technical report, Environmental Defense Fund, URL https://www.edf.org/sites/default/files/Cap_and_trade_auction_FAQ-FINAL_March_1_2013.pdf.
- [76] Paes Leme R, Sivan B, Teng Y (2020) Why do competitive markets converge to first-price auctions? URL <https://arxiv.org/abs/2001.06074>.
- [77] Potfer M, Baudry D, Richard H, Perchet V, Wan C (2024) Improved learning rates in multi-unit uniform price auctions. Globerson A, Mackey L, Belgrave D, Fan A, Paquet U, Tomczak J, Zhang C, eds., *Advances in Neural Information Processing Systems*, volume 37, 130237–130264 (Curran Associates, Inc.), URL https://proceedings.neurips.cc/paper_files/paper/2024/file/eb1fc4b546d293b36622e1925f86f14d-Paper-Conference.pdf.
- [78] Potfer M, Perchet V (2025) Comparing uniform price and discriminatory multi-unit auctions through regret minimization. *Advances in Neural Information Processing Systems*, volume 38, 90154–90196, URL <http://dx.doi.org/10.52202/085713-2716>.
- [79] Rakhlin A, Sridharan K (2014) Online learning with predictable sequences. URL <https://arxiv.org/abs/1208.3728>.
- [80] Regional Greenhouse Gas Initiative (2026) About auctions. <https://www.rggi.org/auctions/about-auctions>, accessed June 5, 2026.
- [81] Regional Greenhouse Gas Initiative (2026) Investments of proceeds. <https://www.rggi.org/investments/proceeds-investments>, accessed June 5, 2026.
- [82] Rottner C (2026) *Procurement Auctions with Variable Quantities*. Ph.D. thesis, University of Bayreuth, URL <https://epub.uni-bayreuth.de/id/eprint/8995/>.
- [83] Rottner C (2026) Watt's better? pay-as-bid vs. uniform pricing in electricity market. *Energy Economics* 157:109284, URL <http://dx.doi.org/10.1016/j.eneco.2026.109284>.
- [84] Schmalensee R, Stavins RN (2017) Lessons learned from three decades of experience with cap and trade. *Review of Environmental Economics and Policy* 11(1):59–79, URL <http://dx.doi.org/10.1093/reep/rew017>.
- [85] Singh N, Vives X (1984) Price and quantity competition in a differentiated duopoly. *The RAND Journal of Economics* 15(4):546–554, ISSN 07416261, URL <http://www.jstor.org/stable/2555525>.
- [86] Stuart D, Hibbard P (2023) The economic impacts of the regional greenhouse gas initiative on ten northeast and mid-atlantic states. Technical report, Analysis Group, URL <https://www.analysisgroup.com/globalassets/insights/publishing/2023-ag-rggi-report.pdf>.
- [87] Sum J, Leung CS, Chang JCC (2025) Analysis and design of a distributed kWTA with application in sealed-bid auctions with bidding price privacy protection. *IEEE Transactions on Neural Networks and Learning Systems* 36(9):16264–16278, URL <http://dx.doi.org/10.1109/TNNLS.2025.3554440>.
- [88] Sum J, Leung CS, Chang JCC (2025) A fast distributed algorithm for uniform price auction with bidding information protection. *Computation* 13(12):294, URL <http://dx.doi.org/10.3390/computation13120294>.
- [89] Syrgkanis V, Agarwal A, Luo H, Schapire RE (2015) Fast convergence of regularized learning in games. URL <https://arxiv.org/abs/1507.00407>.
- [90] Syrgkanis V, Tardos E (2012) Composable and efficient mechanisms. URL <https://arxiv.org/abs/1211.1325>.
- [91] Tierney SF, Schatzki T, Mukerji R (2008) Uniform-pricing versus pay-as-bid in wholesale electricity markets: Does it make a difference? Technical report, Analysis Group and New York Independent System Operator, URL https://www.nyiso.com/documents/20142/2268609/uniformpricing_v_payasbid_tierneyschatzkimukerji_2008.pdf/7b93c886-e3ce-82cc-2dfc-b36ae96efa21.
- [92] Tsakas E (2016) Correlated-belief equilibrium. *Synthese* 193(3):757–779, URL <http://dx.doi.org/10.1007/s11229-015-0791-4>.
- [93] US Environmental Protection Agency (2023) EPA report on the social cost of greenhouse gases: Estimates incorporating recent scientific advances. Technical report, U.S. Environmental Protection Agency, URL https://www.epa.gov/system/files/documents/2023-12/epa_scghg_2023_report_final.pdf.
- [94] Varian HR (2007) Position auctions. *International Journal of Industrial Organization* 25(6):1163–1178, URL <http://dx.doi.org/10.1016/j.ijindorg.2006.10.002>.
- [95] Vlatakis-Gkaragkounis EV, Flokas L, Lianas T, Mertikopoulos P, Piliouras G (2020) No-regret learning and mixed nash equilibria: They do not mix. Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H, eds., *Advances in Neural Information Processing Systems*, volume 33, 1380–1391 (Curran Associates, Inc.), URL https://proceedings.neurips.cc/paper_files/paper/2020/file/0ed9422357395a0d4879191c66f4faa2-Paper.pdf.
- [96] Wang Y, Kong D, Bai Y, Jin C (2022) Learning rationalizable equilibria in multiplayer games. URL <https://arxiv.org/abs/2210.11402>.
- [97] Zhang W, Han Y, Zhou Z, Flores A, Weissman T (2022) Leveraging the hints: Adaptive bidding in repeated first-price auctions. URL <https://arxiv.org/abs/2211.06358>.
- [98] Zhao H, Chen W (2020) Online second price auction with semi-bandit feedback under the non-stationary setting. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(04):6893–6900, URL <http://dx.doi.org/10.1609/aaai.v34i04.6171>.

Appendix. Proofs and Additional Details

A. Algorithm Classes and Initialization/Feedback Robustness

This appendix addresses Sections 2.1, 3, and 7. It formalizes the algorithm-class relationships summarized in Proposition 1, defines the concrete dynamics used in the examples and experiments, and proves the initialization and feedback robustness statements.

A.1. Proof of Proposition 1

This subsection proves the algorithm-class relationships used in Section 3: GBAs strictly refine MBAs, GBAs and no-regret rules are incomparable, and the separable FTRL rules in Proposition 8 are GBAs.

PROPOSITION 1 (Algorithm-class relationships). *GBAs and no-regret algorithms are incomparable (i.e., not subsets of one another), and GBAs form a strict subclass of MBAs. Moreover, every symmetric separable FTRL rule satisfying Proposition 8 is a GBA.*

A.1.1. GBA implies MBA, with an explicit parameter. Fix player n and let C_n be the constant in the GBA upper bound from Definition 4, $p_t^n(a) \leq C_n \gamma_t^n(\Delta_t^n(a))$. Choose any sequence Δ_t with $\Delta_t \gg t^{\alpha_n}$ and $\Delta_t \ll t$; for instance $\Delta_t = t^{(1+\alpha_n)/2}$. The large-gap implication in Definition 4 gives $\gamma_t^n(\Delta_t) \ll 1$. Define

$$\beta_t^n = \sup_{s \geq t} \max \left\{ \sqrt{C_n \gamma_s^n(\Delta_s)}, \Delta_s/s \right\}.$$

Then $\beta_t^n \ll 1$. If $\Delta_t^n(a) > t\beta_t^n$, then for all large t , $\Delta_t^n(a) \geq \Delta_t$ and monotonicity of γ_t^n gives

$$p_t^n(a) \leq C_n \gamma_t^n(\Delta_t^n(a)) \leq C_n \gamma_t^n(\Delta_t) \leq (\beta_t^n)^2 \leq \beta_t^n.$$

Increasing β_t^n on finitely many early periods if needed gives a valid player-specific pruning sequence. Because the game has finitely many players, the common sequence $\beta_t \doteq \max_{n \in [N]} \beta_t^n$ still satisfies $\beta_t \rightarrow 0$ and gives the single MBA parameter required by Definition 5. Thus every GBA is an MBA.

A.1.2. GBA and no-regret are not comparable. We show the incomparability between GBAs and no-regret algorithms using two examples. We first use projected gradient descent (PGD) to show that a no-regret rule need not be a GBA. PGD is a standard no-regret algorithm in full-information repeated games, but it can violate small-gap exploration. Given the current opponent strategy $\mathbf{p}_t^{-n}$, PGD on the simplex updates

$$\mathbf{p}_{t+1}^n = \Pi_{\Delta(\mathcal{A}^n)} \left(\mathbf{p}_t^n + \eta_t (u^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n} \right), \quad \eta_t = \frac{\eta}{\sqrt{t}}. \quad (11)$$

The regret notion here is the same external-regret notion used in the textbook treatment of [17]. Their online-linear-optimization proof applies pathwise to any realized sequence of full-information payoff vectors. In our expected-feedback model, the vector $(u^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n}$ is deterministic once the opponents' mixed strategies at time t are fixed, and the PGD update is deterministic given this vector. Thus the payoff term in Definition 5 is exactly the expected payoff of the learner's mixed action against the deterministic payoff vector at that round, and the Cesa-Bianchi–Lugosi proof methodology carries over without modification.

To show that this is indeed no regret, we borrow the proof from [17]. For any fixed $\mathbf{p}' \in \Delta(\mathcal{A}^n)$, Euclidean projection gives

$$\begin{aligned} \|\mathbf{p}_{t+1}^n - \mathbf{p}'\|_2^2 &\leq \|\mathbf{p}_t^n + \eta_t (u^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n} - \mathbf{p}'\|_2^2 \\ &= \|\mathbf{p}_t^n - \mathbf{p}'\|_2^2 + 2\eta_t \sum_{a \in \mathcal{A}^n} (p_t^n(a) - p'(a)) u^n(a, \mathbf{p}_t^{-n}) + \eta_t^2 \sum_{a \in \mathcal{A}^n} (u^n(a, \mathbf{p}_t^{-n}))^2. \end{aligned}$$

Rearranging gives the following one-round inequality.

$$\sum_{a \in \mathcal{A}^n} (p'(a) - p_t^n(a)) u^n(a, \mathbf{p}_t^{-n}) \leq \frac{\|\mathbf{p}_t^n - \mathbf{p}'\|_2^2 - \|\mathbf{p}_{t+1}^n - \mathbf{p}'\|_2^2}{2\eta_t} + \frac{\eta_t}{2} \sum_{a \in \mathcal{A}^n} (u^n(a, \mathbf{p}_t^{-n}))^2. \quad (12)$$

Because η_t is nonincreasing and the squared Euclidean diameter of the simplex is at most 2, the weighted squared-distance terms satisfy

$$\sum_{t < T} \frac{\|\mathbf{p}_t^n - \mathbf{p}'\|_2^2 - \|\mathbf{p}_{t+1}^n - \mathbf{p}'\|_2^2}{2\eta_t} = \frac{\|\mathbf{p}_1^n - \mathbf{p}'\|_2^2}{2\eta_1} + \sum_{1 < t < T} \frac{\|\mathbf{p}_t^n - \mathbf{p}'\|_2^2}{2} (\eta_t^{-1} - \eta_{t-1}^{-1}) - \frac{\|\mathbf{p}_T^n - \mathbf{p}'\|_2^2}{2\eta_{T-1}} \leq \eta_{T-1}^{-1}. \quad (13)$$

As payoffs are finite, $\sum_{t < T} \eta_t \sum_a (u^n(a, \mathbf{p}_t^{-n}))^2 = O(\sqrt{T})$ when $\eta_t = \eta/\sqrt{t}$. Hence:

$$\left[\sum_{t < T} \sum_{a \in \mathcal{A}^n} (p'(a) - p_t^n(a)) u^n(a, \mathbf{p}_t^{-n}) \right]^+ = O(\sqrt{T}).$$

Summing Equation (12) and using Equation (13) gives the external-regret expression in Definition 5. Under the expected-feedback model in Section 2, no additional expectation over strategy randomization is required:

$$\left[\max_{a \in \mathcal{A}^n} U_T^n(a) - \sum_{t < T} u^n(\mathbf{p}_t) \right]^+ = O(\sqrt{T}) = o(T).$$

The initialization contributes only an $O(1)$ term to this bound. Thus PGD is no-regret in the sense used throughout the paper. However, PGD is not a GBA. Consider the FPA with two bidders sharing the top valuation, as described in Section 5.3.

$$v^1 = v^2 = v \geq 2, \quad b^1, b^2 \in \{0, 1, \dots, v-1\},$$

with ties split evenly and no overbidding. Initialize both players at the weak PNE bid $v - 2$:

$$\mathbf{p}_1^1 = \mathbf{1}_{v-2}, \quad \mathbf{p}_2^1 = \mathbf{1}_{v-2},$$

where $\mathbf{1}_{v-2}$ denotes the point mass that assigns probability one to bid $v - 2$. Against an opponent bidding $v - 2$, each top bidder is indifferent between bidding $v - 2$ and bidding $v - 1$:

$$u^n(v - 2, v - 2) = \frac{v - (v - 2)}{2} = 1, \quad u^n(v - 1, v - 2) = v - (v - 1) = 1.$$

Every lower bid loses and gives payoff 0. Thus $v - 2$ and $v - 1$ are both payoff-maximizing actions at this profile. We now show that PGD remains fixed at the weak PNE. Suppose $\mathbf{p}_t^n = \mathbf{1}_{v-2}$, and suppose $v - 2$ is a payoff-maximizing action against $\mathbf{p}_t^{-n}$.⁷ Then the vector $\mathbf{p}_t^n + \eta_t(u^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n}$ has coordinate $1 + \eta_t u^n(v - 2, \mathbf{p}_t^{-n})$ at $v - 2$, while every other coordinate $a \neq v - 2$ is at most $\eta_t u^n(v - 2, \mathbf{p}_t^{-n})$. For any vector $\mathbf{z}$, Euclidean projection onto the simplex is characterized by

$$\Pi_{\Delta(\mathcal{A}^n)}(\mathbf{z})(a) = [z(a) - \theta]^+, \quad \sum_{a \in \mathcal{A}^n} [z(a) - \theta]^+ = 1, \quad (14)$$

where θ is the unique scalar satisfying the displayed normalization. Applying Equation (14) to the update in Equation (11) with $\theta = \eta_t u^n(v - 2, \mathbf{p}_t^{-n})$ makes the $v - 2$ coordinate one and every other coordinate zero. Thus the projection is exactly $\mathbf{1}_{v-2}$. Hence, by induction,

$$\mathbf{p}_t^1 = \mathbf{1}_{v-2}, \quad \mathbf{p}_t^2 = \mathbf{1}_{v-2} \quad \forall t.$$

But for each bidder n , the unsupported bid $v - 1$ has zero dynamic gap at every time: $\Delta_t^n(v - 1) = 0$, while PGD assigns it zero probability forever: $p_t^n(v - 1) = 0, \forall t$. This violates small-gap exploration. Since $\Delta_t^n(v - 1) = 0$, Definition 4 requires $p_t^n(v - 1) \asymp 1$, while PGD assigns it probability zero at every round. Therefore PGD is no-regret, but it is not a GBA, even in the two-top-bidder FPA.

GBA dynamics need not be no-regret: We now show the opposite direction in a fixed two-player normal-form game. Player 1 has actions A, B , player 2 has actions L, R , player 2's payoff is identically zero, and player 1's payoff is

$$u^1(A, L) = 0, \quad u^1(B, L) = 1, \quad u^1(A, R) = 1, \quad u^1(B, R) = 0.$$

Fix constants $\varepsilon \in (0, 1/2)$ and $\rho \in (0, 1)$ with $\rho + 2\varepsilon(1 - \rho) < 1$. Player 1 uses monotone γ_t^1 with:

$$\gamma_t(0) = 1, \quad \gamma_t(\Delta) = \rho \quad \text{for } 0 < \Delta \leq t^\alpha, \quad \gamma_t(\Delta) = \rho \frac{t^\alpha}{\Delta} \quad \text{for } \Delta > t^\alpha, \quad \text{such that}$$

⁷ Here, the opponent is also initialized at $v - 2$. The same projection calculation applies to both bidders, so the profile is invariant: once both players are at this point, both remain there.

$$p_t^1(a) = \frac{\gamma_t(\Delta_t^1(a))}{\gamma_t(\Delta_t^1(A)) + \gamma_t(\Delta_t^1(B))}, \quad \text{for } a \in \{A, B\}.$$

This is a GBA. Its probabilities are of the same order as the above sensitivity, which is weakly decreasing, remains of constant order below tolerance, and vanishes above tolerance. Since player 2's payoff is always zero, both of player 2's cumulative gaps are identically zero at every history. Let player 2 use the deterministic mixed strategy

$$p_t^2(L) = 1 - \varepsilon, \quad p_t^2(R) = \varepsilon \quad \text{on odd rounds,} \quad p_t^2(L) = \varepsilon, \quad p_t^2(R) = 1 - \varepsilon \quad \text{on even rounds.}$$

Because player 2's gaps are always zero and each action receives probability in $[\varepsilon, 1 - \varepsilon]$, its probabilities are comparable to a constant sensitivity, so Definition 4 holds along this path.

Start player 1 with equal cumulative utilities. On an odd round, player 2 puts probability $1 - \varepsilon$ on L , so player 1's one-period payoff from A is ε and its payoff from B is $1 - \varepsilon$. Thus B gains $1 - 2\varepsilon$ over A . On the following even round, player 2 puts probability $1 - \varepsilon$ on R , so A gains the same amount over B . Therefore player 1's cumulative utilities are tied before every odd round, and B is ahead by exactly $1 - 2\varepsilon$ before every even round.

Consequently, player 1 plays A and B with probability $1/2$ on odd rounds. On even rounds, since $1 - 2\varepsilon \leq t^\alpha$, the gap of A is within tolerance and player 1 plays A with probability $\rho/(1 + \rho)$ and B with probability $1/(1 + \rho)$. Hence player 1's expected payoff in each two-round block is

$$\frac{1}{2} + \frac{\rho(1 - \varepsilon) + \varepsilon}{1 + \rho}.$$

In contrast, each fixed action obtains exactly one unit of payoff in every two-round block: action A gets ε on the odd round and $1 - \varepsilon$ on the even round, while action B gets $1 - \varepsilon$ on the odd round and ε on the even round. The regret per two-round block is therefore

$$1 - \frac{1}{2} - \frac{\rho(1 - \varepsilon) + \varepsilon}{1 + \rho} = \frac{1 - \rho - 2\varepsilon(1 - \rho)}{2(1 + \rho)} > 0.$$

Thus player 1's external regret grows linearly. To be pedantic, we note that our example uses deterministic learning rules that output mixed strategies under the expected-feedback model. As the game is static, the dynamics are fully determined from the initialization, and do not use an arbitrary, stochastic payoff sequence nor do they require any non-monotone dependence of probability on the gap. The conclusion is that the GBA property alone does not imply no-regret. $\square$

A.1.3. A sufficient condition for symmetric separable FTRL to be gap-based The following proposition gives a simple sufficient condition under which a symmetric separable FTRL rule is a GBA. Symmetry means that the same scalar regularizer is applied to every action.

PROPOSITION 8 (Symmetric separable FTRL is gap-based). *Let $\mathcal{A}$ be a finite action set, let $\alpha \in (0, 1)$, and define $\Psi(\mathbf{x}) \doteq \sum_{a \in \mathcal{A}} \psi(x(a))$. Consider the FTRL rule*

$$\mathbf{p}_t \in \arg \max_{\mathbf{x} \in \Delta(\mathcal{A})} \left\{ \sum_{a \in \mathcal{A}} x(a) U_t(a) - t^\alpha \Psi(\mathbf{x}) \right\}. \quad (15)$$

Here $\psi : (0, 1] \rightarrow \mathbb{R}$ is twice continuously differentiable. Assume that, for some constant $m > 0$,

$$x\psi''(x) \geq m \quad \text{for every } x \in (0, 1]. \quad (16)$$

For $\Delta \geq 0$, let $\gamma_t(\Delta)$ be the unique $x \in (0, 1]$ satisfying $\psi'(1) - \psi'(x) = \Delta/t^\alpha$. Then γ_t is weakly decreasing, and $\exists c > 0$ depending only on ψ and $|\mathcal{A}|$, such that, for every t and $a \in \mathcal{A}$,

$$c \gamma_t(\Delta_t(a)) \leq p_t(a) \leq \gamma_t(\Delta_t(a)). \quad (17)$$

Consequently, this FTRL rule is a GBA with tolerance exponent α .

Proof. We divide the argument into four steps. First, we show that the second derivative condition implies steepness and also invertibility. Second, we prove that the optimizer exists, is unique, and lies strictly within the probability simplex interior (i.e., all positive probabilities) due to the steepness condition. Once we have this, the KKT conditions verify optimality of this interior solution, and we can express these probabilities and their decay rate using the sensitivity functions and gaps.

Step 1: curvature gives steepness and an inverse. Condition (16) implies that ψ is strictly convex. Moreover,

$$\psi'(1) - \psi'(x) = \int_x^1 \psi''(s) ds \geq m \int_x^1 \frac{ds}{s} = m \log\left(\frac{1}{x}\right) \rightarrow \infty \quad \text{as } x \rightarrow 0. \quad (18)$$

Thus $\psi'(x) \rightarrow -\infty$ as $x \rightarrow 0$. The map $x \mapsto \psi'(1) - \psi'(x)$ is continuous and strictly decreasing from $+\infty$ to 0 on $(0, 1]$, so the definition of γ_t in Proposition 8 is well posed and γ_t is weakly decreasing in Δ .

Step 2: the optimizer is unique and interior. The objective in Equation (15) is strictly concave, so it has at most one maximizer. Compactness of the simplex and the usual lower-semicontinuous

extension of ψ to the boundary give existence. Equation (18) implies $\psi'(0+) = -\infty$: if a maximizer assigned zero probability to action a and positive probability to action b , moving an arbitrarily small amount of probability from b to a would have positive one-sided derivative. Hence the unique maximizer is interior.

At an interior optimizer, the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient because the objective is concave and the simplex constraint is affine. Therefore, for some multiplier $\lambda_t \in \mathbb{R}$,

$$U_t(a) - t^\alpha \psi'(p_t(a)) = \lambda_t \quad \forall a \in \mathcal{A}. \quad (19)$$

Step 3: the KKT system gives the upper bound. Choose $a^\star \in \arg \max_{b \in \mathcal{A}} U_t(b)$. Subtracting Equation (19) for a from the same equation for $a^\star$ gives

$$\frac{\Delta_t(a)}{t^\alpha} = \psi'(p_t(a^\star)) - \psi'(p_t(a)) \geq 0. \quad (20)$$

Since ψ' is strictly increasing, $p_t(a^\star) \geq p_t(a)$ for every a , and therefore $p_t(a^\star) \geq 1/|\mathcal{A}|$. Define $\theta_t \doteq \psi'(1) - \psi'(p_t(a^\star))$. Then $0 \leq \theta_t \leq \psi'(1) - \psi'(1/|\mathcal{A}|)$, and Equation (20) becomes

$$\psi'(1) - \psi'(p_t(a)) = \frac{\Delta_t(a)}{t^\alpha} + \theta_t. \quad (21)$$

The right-hand side is at least $\Delta_t(a)/t^\alpha$. By the definition of γ_t and monotonicity of $x \mapsto \psi'(1) - \psi'(x)$, Equation (21) implies $p_t(a) \leq \gamma_t(\Delta_t(a))$.

Step 4: the normalization shift changes probability by only a constant factor. Subtracting the defining equality for $\gamma_t(\Delta_t(a))$ from Equation (21) gives

$$\theta_t = \psi'(\gamma_t(\Delta_t(a))) - \psi'(p_t(a)) = \int_{p_t(a)}^{\gamma_t(\Delta_t(a))} \psi''(s) ds. \quad (22)$$

Using Condition (16) in Equation (22),

$$\theta_t \geq m \log \left(\frac{\gamma_t(\Delta_t(a))}{p_t(a)} \right), \quad p_t(a) \geq e^{-\theta_t/m} \gamma_t(\Delta_t(a)).$$

Because $\theta_t \leq \psi'(1) - \psi'(1/|\mathcal{A}|)$, the constant $c \doteq \exp(-[\psi'(1) - \psi'(1/|\mathcal{A}|)]/m) > 0$ gives $p_t(a) \geq c \gamma_t(\Delta_t(a))$. Together with the upper bound from Equation (21), this proves Equation (17).

Finally, if $\Delta_t(a) \lesssim t^\alpha$, the argument of the inverse defining γ_t remains bounded, so $\gamma_t(\Delta_t(a))$ is bounded away from zero. If $\Delta_t(a) \gg t^\alpha$, Equation (18) gives $\gamma_t(\Delta_t(a)) \rightarrow 0$. Equation (17) therefore verifies the small-gap and large-gap requirements in Equations (Small Gap Exploration) and (Large Gap Pruning), together with the probability comparison in Equation (7). Thus the FTRL rule is a GBA. $\square$

Why the assumption is natural. Condition (16) has two roles. It makes the regularizer steep enough near zero that every action keeps positive probability, and it ensures that the bounded normalization term created by the simplex constraint changes probabilities by at most a constant factor. Table 4 verifies the condition for the interior FTRL regularizers used in this paper.

Algorithm / FTRL rule	Regularizer $\psi(x)$	$x\psi''(x)$	Implication
Hedge, multiplicative weights, or exponential weights	$x \log x$	1	The condition holds with $m = 1$.
Tsallis FTRL, $q_{ts} \in (0, 1)$	$\frac{x - x^{q_{ts}}}{q_{ts}(1 - q_{ts})}$	$x^{q_{ts}-1} \geq 1$	The condition holds with $m = 1$. Its tail has the same polynomial order as polynomial weights with exponent $\beta = 1/(1 - q_{ts})$.
Log-barrier FTRL	$-\log x$	$1/x \geq 1$	The condition holds with $m = 1$. Its tail has the same polynomial order as polynomial weights with exponent $\beta = 1$.
Power-barrier FTRL, $r > 0$	$\frac{x^{-r}}{r}$	$(r + 1)x^{-r-1} \geq r + 1$	The condition holds with $m = r + 1$.
Euclidean FTRL (non-GBA)	$x^2/2$	x	The condition fails near zero. Euclidean FTRL may assign zero probability to actions whose gaps remain at the tolerance scale and is not covered by the proposition.

Table 4 Regularizers corresponding to the FTRL algorithms used in the paper. The first four rows are steep and satisfy Proposition 8. The shaded Euclidean FTRL row is included as a non-steep boundary case. The Tsallis and polynomial-weights updates have the same tail order when $\beta = 1/(1 - q_{ts})$; this does not assert that their normalized updates are identical.

Scope of the condition. The curvature condition is sufficient but not necessary. The proof only requires the regularizer to keep the FTRL optimizer in the simplex interior and to ensure that bounded additive shifts in the first-order condition change probabilities by at most constant factors. Other regularizers may satisfy these two properties without satisfying Condition (16) uniformly, and action-dependent regularizers can also be allowed when their behavior near zero is comparable. We omit these extensions because every FTRL rule used in this paper satisfies Condition (16).

A.2. Algorithmic classification

This subsection provides details regarding the algorithm list in Section 3 and the experiment protocol in Section 6.1. It shows which simulated dynamics are genuine GBAs, which are comparison dynamics, and when the GBA families used in the paper satisfy forgiveness.

We first define the algorithm families used in Sections 3 and 6: symmetric separable FTRL rules, including Hedge and Tsallis FTRL, polynomial weights, and ϵ_t -buffered GBAs. We also describe Euclidean FTRL, FTL, and PGD as comparison dynamics that are not GBAs under Definition 4.

- (i) **Separable FTRL rules, including Hedge.** The symmetric separable FTRL rules covered by Proposition 8 use the regularization scale t^{α_n} :

$$\mathbf{p}_t^n \in \arg \max_{\mathbf{x} \in \Delta(\mathcal{A}^n)} \left\{ \sum_{a \in \mathcal{A}^n} x(a) U_t^n(a) - t^{\alpha_n} \Psi^n(\mathbf{x}) \right\}, \quad \Psi^n(\mathbf{x}) = \sum_{a \in \mathcal{A}^n} \psi^n(x(a)).$$

Here α_n is the GBA tolerance exponent from Definition 4. Hedge, also known as multiplicative weights or exponential weights, is the entropy-FTRL special case:

$$p_t^n(a) = \frac{\exp(t^{-\alpha_n} U_t^n(a))}{\sum_{a' \in \mathcal{A}^n} \exp(t^{-\alpha_n} U_t^n(a'))} = \frac{\exp(-t^{-\alpha_n} \Delta_t^n(a))}{\sum_{a' \in \mathcal{A}^n} \exp(-t^{-\alpha_n} \Delta_t^n(a'))}.$$

Tsallis, log-barrier, and power-barrier FTRL also satisfy Proposition 8. For Tsallis FTRL with regularizer parameter $q_{ts} \in (0, 1)$,

$$\Psi_{q_{ts}}(\mathbf{p}) = \frac{1}{q_{ts}(1 - q_{ts})} \sum_{a \in \mathcal{A}^n} (p(a) - p(a)^{q_{ts}}), \quad p_t^n(a) \propto (1 + t^{-\alpha_n} \Delta_t^n(a))^{-1/(1 - q_{ts})}.$$

Thus its polynomial-tail exponent is $\beta_n = 1/(1 - q_{ts})$.

- (ii) **Polynomial weights.** The polynomial-weights rule uses the same tolerance exponent α_n and a separate tail exponent $\beta_n > 0$:

$$p_t^n(a) = \frac{(1 + t^{-\alpha_n} \Delta_t^n(a))^{-\beta_n}}{\sum_{a' \in \mathcal{A}^n} (1 + t^{-\alpha_n} \Delta_t^n(a'))^{-\beta_n}}.$$

This rule is a GBA because its probability is a decreasing function of the cumulative gap, but it is not generally FTRL-based or no-regret. The common exponent β_n lets us state the forgiveness condition for polynomial weights and Tsallis FTRL in the same form.

- (iii) **ϵ_t -buffered GBAs.** Given a base GBA $\mathbf{p}_{t,0}^n$ and a vanishing exploration floor $\epsilon_t^n \rightarrow 0$, define

$$p_t^n(a) = (1 - \epsilon_t^n) p_{t,0}^n(a) + \frac{\epsilon_t^n}{|\mathcal{A}^n|}.$$

For example, buffered polynomial weights is given by

$$p_{t,0}^n(a) \propto (1 + t^{-\alpha_n} \Delta_t^n(a))^{-\beta_n}, \quad \epsilon_t^n = t^{-s_n},$$

where $s_n > 0$ is the exploration-floor decay exponent. This notation is distinct from the Tsallis parameter q_{ts} . A linear-gap action then has probability of order $t^{-\min\{\beta_n(1-\alpha_n), s_n\}}$. The floor preserves the GBA property, but it can destroy forgiveness by making linear-gap actions occur too frequently.

(iv) **Euclidean FTRL, FTL, and projected gradient / mirror descent.** For t^{α_n} regularization, Euclidean FTRL solves

$$\mathbf{p}_t^n \in \arg \max_{\mathbf{p} \in \Delta(\mathcal{A}^n)} \left\{ \sum_a p(a) U_t^n(a) - \frac{t^{\alpha_n}}{2} \sum_a p(a)^2 \right\}.$$

FTL chooses a cumulative-utility maximizer,

$$a_t^n \in \arg \max_{a \in \mathcal{A}^n} U_t^n(a),$$

and assigns probability one to a_t^n . It is an MBA because any action with a large positive gap is never selected after it falls behind the leader, but it is not a GBA because it can assign zero probability to another zero-gap action. PGD on the simplex for payoff maximization updates

$$\mathbf{p}_{t+1}^n = \Pi_{\Delta(\mathcal{A}^n)} (\mathbf{p}_t^n + \eta_t \mathbf{u}_t^n),$$

where $\mathbf{u}_t^n = (u_t^n(a, \mathbf{p}_t^{-n}))_{a \in \mathcal{A}^n}$ is the full-information payoff vector and η_t is a standard no-regret step size. Euclidean FTRL and PGD are also not GBAs under Definition 4. Euclidean FTRL can assign zero probability to an action whose gap remains of the same order as the tolerance. For PGD, consider two actions a, b , initialize at the vertex $p_1(a) = 1, p_1(b) = 0$, and give both actions payoff zero at every round. PGD remains at the same vertex forever and has zero regret. Both actions have zero cumulative gap, so small-gap exploration requires $p_t(b) \asymp 1$, whereas PGD assigns $p_t(b) = 0$.

PROPOSITION 9 (Algorithmic conditions for forgiveness). *For the algorithm families defined in items (i)–(iii) of Appendix A.2:*

(i) *Hedge is forgiving.*

(ii) *Polynomial weights, Tsallis FTRL, and log-barrier FTRL are n -forgiving if and only if*

$$\beta_i(1 - \alpha_i) \geq 1 - \alpha_n \quad \text{for every } i \neq n,$$

where $\beta_i = 1/(1 - q_{ts})$ for a Tsallis rule with parameter q_{ts} and $\beta_i = 1$ for log-barrier FTRL. In particular, they are forgiving when $\beta_i(1 - \alpha_i) \geq 1$ for every player i , and Tsallis and log-barrier FTRL are forgiving when all players share a tolerance exponent.

(iii) *Buffered polynomial weights with $\epsilon_t^i = t^{-s_i}$ are n -forgiving if and only if*

$$\min\{\beta_i(1 - \alpha_i), s_i\} \geq 1 - \alpha_n \quad \text{for every } i \neq n.$$

Proof.

Hedge. For a linear gap κt , Hedge assigns sensitivity $\gamma_t^i(\kappa t) = \exp(-\kappa t^{1-\alpha_i})$, whose cumulative sum is bounded. Hence it is n -forgiving for every n .

Polynomial weights, Tsallis FTRL, and log-barrier FTRL. Put $r_i = \beta_i(1 - \alpha_i)$. A linear gap has sensitivity of order t^{-r_i} , whose cumulative sum is of order t^{1-r_i} when $r_i < 1$, logarithmic when $r_i = 1$, and bounded when $r_i > 1$. Since $\alpha_n > 0$, this is $O(t^{\alpha_n})$ exactly when $r_i \geq 1 - \alpha_n$. By Proposition 8, Tsallis FTRL has the same probability order with $\beta_i = 1/(1 - q_{ts})$, and log-barrier FTRL has sensitivity $(1 + \Delta/t^{\alpha_i})^{-1}$, i.e., $\beta_i = 1$.

Buffered polynomial weights. A linear-gap action has probability of order $t^{-\min\{\beta_i(1-\alpha_i), s_i\}}$. Summing this sequence gives the stated necessary and sufficient condition. $\square$

A.3. Initialization and Feedback Robustness

This Appendix pertains to Section 7. It proves two robustness facts used in the paper: Proposition 10 shows that finite initial histories change cumulative gaps by only a bounded amount, and Proposition 11 gives high-probability static-gap comparisons under realized full-information feedback. The realized-feedback extension is used only when every tolerance exponent satisfies $\alpha_n > 1/2$.

PROPOSITION 10 (Robustness to finite priors and time shifts). *Fix a finite previous history of length T and finite shifted initialization values $\bar{U}_0^n(a)$. Define*

$$\bar{U}_t^n(a) \doteq \bar{U}_0^n(a) + \sum_{\tau=T}^{T+t-1} u^n(a, \mathbf{p}_\tau^{-n}), \quad \bar{\Delta}_t^n(a) \doteq \max_b \bar{U}_t^n(b) - \bar{U}_t^n(a).$$

Let $\Delta_{T+t}^n(a)$ denote the unshifted gap computed from the same infinite path starting at time 0. Then, for every player n and action a ,

$$|\bar{\Delta}_t^n(a) - \Delta_{T+t}^n(a)| \lesssim 1, \quad (t+T)^{\alpha_n} \asymp t^{\alpha_n}.$$

Proof. Let $C^n(a) = \bar{U}_0^n(a) - U_T^n(a)$. The previous history and the shifted initialization vector are finite, so $\max_a |C^n(a)| < \infty$. Moreover, $\bar{U}_t^n(a) = U_{T+t}^n(a) + C^n(a)$ for every action a . Hence, for any two actions a, b ,

$$|[\bar{U}_t^n(b) - \bar{U}_t^n(a)] - [U_{T+t}^n(b) - U_{T+t}^n(a)]| \leq 2 \max_c |C^n(c)|.$$

Taking maxima over the finite action set gives $|\bar{\Delta}_t^n(a) - \Delta_{T+t}^n(a)| \lesssim 1$. The time-scale identity $(t+T)^{\alpha_n} \asymp t^{\alpha_n}$ is immediate because T is fixed and $\alpha_n \in (0, 1)$. $\square$

Before stating the realized-feedback result, fix a candidate limiting distribution $\mathbf{p}$. We say the empirical opponent distributions are ε -close to $\mathbf{p}$ at time t if, for every player n and action $a \in \mathcal{A}^n$,

$$|u^n(a, \bar{\mathbf{p}}_t^{-n}) - u^n(a, \mathbf{p}^{-n})| \leq \varepsilon/2.$$

The next proposition shows that, when the empirical opponent distributions are close to $\mathbf{p}$, realized cumulative suboptimality concentrates around the corresponding static gaps times t .

PROPOSITION 11 (Realized-feedback static-gap bounds). *Assume payoffs lie in $[0, \bar{u}]$. At each round t , suppose the realized-feedback learner observes full-information counterfactual payoffs $\hat{u}_t^n(a)$ satisfying*

$$\mathbb{E}[\hat{u}_t^n(a) \mid \mathcal{F}_{t-1}] = u^n(a, \mathbf{p}_t^{-n}), \quad \hat{u}_t^n(a) \in [0, \bar{u}]$$

where the expectation is over the round- t realization randomness, conditional on the past. Let

$$\hat{U}_t^n(a) \doteq \sum_{\tau < t} \hat{u}_\tau^n(a), \quad \hat{\Delta}_t^n(a) \doteq \max_{b \in \mathcal{A}^n} \hat{U}_t^n(b) - \hat{U}_t^n(a)$$

denote the realized feedback analogues of the cumulative rewards and gaps, $U_t^n(a)$ and $\Delta_t^n(a)$ respectively. Let $K \doteq \sum_n |\mathcal{A}^n|$. Fix $t \geq 2$, $\varepsilon > 0$, and a candidate distribution $\mathbf{p}$. With probability at least $1 - 2K/t^2$, if the empirical opponent distributions are ε -close to $\mathbf{p}$ at time t , then, for every player n and action $a \in \mathcal{A}^n$,

$$[\delta^n(a; \mathbf{p}) - \varepsilon]^+ t - 4\bar{u}\sqrt{t \log t} \leq \hat{\Delta}_t^n(a) \leq (\delta^n(a; \mathbf{p}) + \varepsilon)t + 4\bar{u}\sqrt{t \log t}.$$

Here $\bar{u}$ is any uniform bound on payoffs. In the multi-unit auctions of Section 6, no overbidding keeps each bidder's payoff between zero and the total value of the units it wins, so whenever every marginal value is at most 1, we may take $\bar{u} = M$.

Proof. To prove this, we first fix a player n and action a . Define

$$Z_t^n(a) \doteq \sum_{\tau < t} (\hat{u}_\tau^n(a) - u^n(a, \mathbf{p}_\tau^{-n}))$$

to be the de-meaned rewards. Because $\mathbf{p}_\tau$ is determined before the round- τ realization is drawn, $(Z_t^n(a))_t$ is a martingale with increments in $[-\bar{u}, \bar{u}]$. Hoeffding–Azuma gives

$$\mathbb{P}\left(|Z_t^n(a)| > 2\bar{u}\sqrt{t \log t}\right) \leq 2t^{-2}.$$

Taking a union bound over the K player-action pairs, with probability at least $1 - 2K/t^2$,

$$\max_{n,a} |Z_t^n(a)| \leq 2\bar{u}\sqrt{t \log t}.$$

On this event, for every player n and pair of actions a, b ,

$$\begin{aligned}\widehat{U}_t^n(b) - \widehat{U}_t^n(a) &= \sum_{\tau < t} (u^n(b, \mathbf{p}_\tau^{-n}) - u^n(a, \mathbf{p}_\tau^{-n})) + Z_t^n(b) - Z_t^n(a) \\ &= t(u^n(b, \bar{\mathbf{p}}_t^{-n}) - u^n(a, \bar{\mathbf{p}}_t^{-n})) + Z_t^n(b) - Z_t^n(a),\end{aligned}$$

where the second equality uses linearity of expected payoff in the opponent distribution. Since $|Z_t^n(b) - Z_t^n(a)| \leq 4\bar{u}\sqrt{t \log t}$, taking the maximum over $b \in \mathcal{A}^n$ gives

$$\widehat{\Delta}_t^n(a) = t \left[\max_{b \in \mathcal{A}^n} u^n(b, \bar{\mathbf{p}}_t^{-n}) - u^n(a, \bar{\mathbf{p}}_t^{-n}) \right] \pm 4\bar{u}\sqrt{t \log t}.$$

If the empirical opponent distributions are ε -close to $\mathbf{p}$, then every pairwise payoff difference is within ε of its value under $\mathbf{p}^{-n}$. Hence the bracketed term is between $[\delta^n(a; \mathbf{p}) - \varepsilon]^+$ and $\delta^n(a; \mathbf{p}) + \varepsilon$, which proves the claim. $\square$

COROLLARY 3 (Realized-feedback extension). *Suppose every tolerance exponent satisfies $\alpha_n > 1/2$, and suppose the players' GBA conditions in Definition 4 are imposed on the realized gaps $\widehat{\Delta}_t^n$. Under the realized full-information feedback model of Proposition 11, the conclusions of Theorems 1, 3, and 4 hold almost surely.*

Proof. The failure probability in Proposition 11 is $2K/t^2$, and $\sum_{t \geq 2} 2K/t^2 < \infty$. The Borel–Cantelli lemma therefore implies that, almost surely, for all sufficiently large t and every player-action pair,

$$|\widehat{U}_t^n(a) - U_t^n(a)| = O(\sqrt{t \log t}), \quad |\widehat{\Delta}_t^n(a) - \Delta_t^n(a)| = O(\sqrt{t \log t}).$$

Because $\alpha_n > 1/2$, both errors are $o(t^{\alpha_n})$. A positive static gap still gives a linear realized gap, so the proof of Theorem 1 and the positive-gap step in Theorem 4 are unchanged. In the proof of Theorem 3, Equation (42) acquires only an additional $O(\sqrt{t \log t}) = o(t^{\alpha_n})$ term, while Equation (41) is stated directly for the realized GBA gap. Hence the substitute remains within tolerance on the same linear number of rounds, and the rest of the proof is unchanged. $\square$

B. Large-Gap Support-Refinement Proofs

This Appendix supports the large-gap results in Section 4. We first prove the static-gap bound used in Theorem 1. We then give the two-player support result that explains why BR-CCE optimization contains NE support selection as a special case, followed by the PAB welfare proof.

B.1. Static-gap bounds under convergence

The following proposition converts a static payoff gap at a candidate limit into a statement about cumulative dynamic gaps. It is used in both Theorem 1 and the forgiving proof.

PROPOSITION 12 (Static gap bounds under convergence). Suppose the empirical opponent distributions satisfy $\bar{\mathbf{p}}_t^{-n} \rightarrow \mathbf{p}^{-n}$. Then, for every player n , action $a \in \mathcal{A}^n$, and $\varepsilon > 0$, all sufficiently large t satisfy

$$[\delta^n(a; \mathbf{p}) - \varepsilon]^+ t \leq \Delta_t^n(a) \leq (\delta^n(a; \mathbf{p}) + \varepsilon)t. \quad (23)$$

Consequently, dividing both sides by t and sending $\varepsilon \rightarrow 0$

$$\frac{\Delta_t^n(a)}{t} \rightarrow \delta^n(a; \mathbf{p}). \quad (24)$$

If the dynamics converge to $\mathbf{p}$, then the empirical opponent marginals also converge, so Equation (24) applies. In particular, positive-static-gap actions have linear dynamic gap, while every zero-static-gap action has $o(t)$ dynamic gap.

Proof. Fix player n and action a . From the cumulative-utility definition in Equation (3) and the empirical opponent distribution in Equation (5), every comparison action $b \in \mathcal{A}^n$ satisfies

$$\frac{U_t^n(b) - U_t^n(a)}{t} = u^n(b, \bar{\mathbf{p}}_t^{-n}) - u^n(a, \bar{\mathbf{p}}_t^{-n}) + O(1/t), \quad (25)$$

where the $O(1/t)$ term comes only from finite initialization. Since the game is finite and payoffs are continuous in the opponent distribution, $\bar{\mathbf{p}}_t^{-n} \rightarrow \mathbf{p}^{-n}$ makes the right-hand side of Equation (25) converge uniformly over b to $u^n(b, \mathbf{p}^{-n}) - u^n(a, \mathbf{p}^{-n})$. Taking the maximum over b , and using the definitions of the dynamic and static gaps in Equations (4) and (2), proves Equation (24). Equation (23) follows immediately. The final claims follow from Definitions 1 and 2. $\square$

Frequency consequences. If $\delta^n(a; \mathbf{p}) > 0$, monotonicity of γ_t^n , Equation (23), and the probability comparison in Equation (7) imply, for every sufficiently small $\varepsilon > 0$,

$$P_t^n(a) \lesssim 1 + \sum_{\tau < t} \gamma_\tau^n((\delta^n(a; \mathbf{p}) - \varepsilon)\tau), \quad P_t^n(a) \gtrsim \sum_{\tau < t} \gamma_\tau^n((\delta^n(a; \mathbf{p}) + \varepsilon)\tau) \quad (26)$$

The corresponding probability comparisons follow by applying the same monotonicity argument before summing. The additive constant in the first inequality accounts only for finitely many early rounds; the second inequality is used only for rate comparisons. If $\delta^n(a; \mathbf{p}) = 0$, Equation (24) gives $\Delta_t^n(a) = o(t)$.

B.2. Two-player support collapse

COROLLARY 4 (Two-player support collapse). *If $N = 2$ and $\mathbf{p} \in \mathcal{P}^*$, then the marginal profile $(\mathbf{p}^1, \mathbf{p}^2)$ is an MNE. Consequently, any empirical limit of GBA dynamics in a two-player game has an MNE marginal profile, even without no regret.*

Proof. With two players, each opponent joint marginal is simply the other player's marginal. Let $\mathbf{p} \in \mathcal{P}^*$. By the definition of $\mathcal{P}^*$ in Proposition 2, for every $a \in \text{Supp}(\mathbf{p}^1)$ and $a' \in \text{Supp}(\mathbf{p}^2)$,

$$a \in \arg \max_{b \in \mathcal{A}^1} u^1(b, \mathbf{p}^2), \quad a' \in \arg \max_{b \in \mathcal{A}^2} u^2(b, \mathbf{p}^1).$$

This is exactly the MNE support condition in Definition 1. If GBA dynamics have an empirical limit, Theorem 1 places that limit in $\mathcal{P}^*$, so its marginals form an MNE. $\square$

Because Corollary 4 embeds the usual two-player NE support-selection problem already inside $\mathcal{P}^*$, optimizing over BR-CCE—which additionally imposes the CCE inequalities—contains this support-selection difficulty. This is why our computations use support enumeration or mixed-integer support formulations rather than a single linear program.

B.3. Welfare Guarantees for Section 4.1

We now prove the PAB welfare hierarchy in Theorem 2. Lemma 1 first removes the two-dimensional bid-vector complication: either the supported outcomes are already efficient, or every supported outcome can be represented by one scalar capped-threshold bid per bidder. We then use support optimality to determine the resulting marginal bid distributions, obtains the 11/14 guarantee for arbitrary support-optimal $\mathbf{p} \in \mathcal{P}^*$, and finally sharpens the bound to 5/6 when including the CCE inequalities. Separate examples establish tightness, zero welfare under CR support alone, and the strict CCE_{CR} –BR–CCE separation.

LEMMA 1 (Reduction to capped bids). *Let $N = M = 2$ and let $\mathbf{p} \in \mathcal{P}^*$ for PAB as in Section 4.1. Either (i) every supported allocation is welfare efficient, or (ii) each supported bid $\mathbf{b}^n$ can be mapped to the capped-threshold bid $\widehat{\mathbf{b}}^n = (b_1^n, \min\{b_1^n, v_2^n - 1\})$ without changing any supported allocation, payment, bidder utility, or welfare. For each supported profile $\mathbf{b}$, let $\widehat{\mathbf{b}}$ denote its transformed profile, and let $\widehat{\mathbf{p}}$ denote the distribution of $\widehat{\mathbf{b}}$ when $\mathbf{b} \sim \mathbf{p}$. In the second case, $\text{WEL}(\widehat{\mathbf{p}}; \mathbf{v}) = \text{WEL}(\mathbf{p}; \mathbf{v})$. Moreover, $\widehat{\mathbf{p}} \in \mathcal{P}^*$; if $\mathbf{p}$ is a CCE, then $\widehat{\mathbf{p}}$ is a CCE; and if $\mathbf{p}$ is a product distribution, then $\widehat{\mathbf{p}}$ is also a product distribution. Consequently, an MNE is mapped to an MNE.*

Proof. Our proof has several parts. Part 1 handles the zero-utility case and proves that every supported allocation must already be efficient. We introduce minimum supported bids in Part 2, after restricting attention to the positive-utility case, because they identify the smallest possible bids required for allocation and thus show that any smaller bids will go unallocated throughout the support. Parts 3 and 4 construct the capped-threshold maps under the two possible orderings of the bidders' second values. Part 5 verifies that the maps preserve each agent's payments, allocations, and utilities whilst weakly decreasing each bidder's incentive to deviate. This consequently also preserves the original solution class of $\mathbf{p}$.

Part 1: The zero-utility case. We first consider the case when one bidder n has 0 expected utility under $\mathbf{p} \in \mathcal{P}^*$. By support-optimality, they must have best-response utility zero, and thus, every supported bid of bidder n then has expected utility zero. This implies that it is impossible for them to win any units under any bid profile $\mathbf{b}^n$ without overbidding—otherwise, they would be able to profitably deviate. Hence bidder n cannot win at any supported profile.

If some supported allocation were welfare inefficient, one of bidder n 's losing marginal values would strictly exceed the value of a winning unit. Let β be the bid attached to that lower-valued winning unit. Bidder n can bid the smallest integer that defeats β under the fixed tie-breaking rule on the relevant marginal units. This deviation is feasible because the displaced winning value is strictly below bidder n 's marginal value, and it earns strictly positive surplus whenever the selected supported opponent profile occurs while never giving negative utility elsewhere. This contradicts zero best-response utility. Therefore every supported allocation is efficient, which is statement (i) in the lemma.

Part 2: Computing the smallest bids required for allocation. We henceforth assume that both bidders have positive best-response utility. We first prove a useful property that allows us to decouple the expected utility of a bid profile $\mathbf{b}^n$ as a sum of the expected utilities coming from both of its bids b_1^n, b_2^n . In particular, expected utility is separable across the two bid coordinates:

$$u^1(\mathbf{b}^1, \mathbf{p}^2) = (v_1^1 - b_1^1) \Pr(b_2^2 \leq b_1^1 \mid \mathbf{b}^2 \sim \mathbf{p}^2) + (v_2^1 - b_2^1) \Pr(b_1^2 \leq b_2^1 \mid \mathbf{b}^2 \sim \mathbf{p}^2). \quad (27)$$

and bidder 2 has the analogous expression with strict inequalities because bidder 1 wins ties. The feasible bid set is closed under coordinatewise minimum and maximum, so for any two feasible bids $\mathbf{b}'$ and $\mathbf{b}''$,

$$u^n(\mathbf{b}', \mathbf{p}^{-n}) + u^n(\mathbf{b}'', \mathbf{p}^{-n}) = u^n(\mathbf{b}' \wedge \mathbf{b}'', \mathbf{p}^{-n}) + u^n(\mathbf{b}' \vee \mathbf{b}'', \mathbf{p}^{-n}).$$

If $\mathbf{b}'$ and $\mathbf{b}''$ are best responses, both terms on the right are at most the best-response utility and their sum is twice that utility. Both are therefore best responses. Iterating over the finite support shows that the coordinatewise minima $\underline{\mathbf{b}}^n$ is a best response for bidder n where:

$$\underline{\mathbf{b}}^n \doteq (\underline{b}_1^n, \underline{b}_2^n), \quad \underline{b}_m^n \doteq \min_{\mathbf{b}^n \in \text{Supp}(\mathbf{p}^n)} b_m^n.$$

Because this best response has positive utility, $\underline{\mathbf{b}}^n$ wins a unit against some supported opponent bid. Replacing that opponent bid by its coordinatewise minimum can only lower the competing bids, so each bidder wins at least one unit at $(\underline{\mathbf{b}}^1, \underline{\mathbf{b}}^2)$. The profile therefore splits the two units. Each minimum second bid is losing at this profile, and it remains losing against every other supported opponent bid because all such bids are coordinatewise weakly larger. Thus, replacing the minimum second coordinate by zero preserves utility, so $(\underline{b}_1^n, 0)$ is also a best response for each bidder.

Part 3: The case of $v_2^1 < v_2^2$. Part 2 identified the smallest bids required for allocation. We now use those identities to define the transformed profile $\widehat{\mathbf{b}}$ and show that each supported profile $\mathbf{b}$ and its image $\widehat{\mathbf{b}}$ have the same allocation, payments, and bidder utilities. Under this assumption of $v_2^1 < v_2^2$, every bidder-1 second bid is at most $v_2^1 - 1$. If $\underline{b}_1^2 > v_2^1$, bidder 2 can lower the first coordinate of the best response $(\underline{b}_1^2, 0)$ to v_2^1 , preserve its baseline unit, and pay less. Hence $\underline{b}_1^2 \leq v_2^1$.

Suppose $\underline{b}_1^2 < v_2^1$. At the split minimum profile, if $\underline{b}_1^1 \geq \underline{b}_1^2$, bidder 1 can raise its losing second bid to $\underline{b}_1^2$, win the tie, and obtain a profitable second unit. If $\underline{b}_1^1 < \underline{b}_1^2$, bidder 2 can instead raise its losing second bid to $\underline{b}_1^1 + 1$ and profitably win the second unit. Both cases contradict best-response optimality. Thus $\underline{b}_1^2 = v_2^1$, and the same second-unit deviation gives $\underline{b}_1^1 \geq v_2^1$.

Now fix a supported bidder-2 bid (b_1^2, b_2^2) . Lowering its first coordinate to $\max\{v_2^1, b_2^2\}$ preserves its baseline unit. If the original action won both units, its second bid already beat bidder 1's first bid, so this change also preserves the second unit. Best-response optimality therefore requires $b_1^2 = \max\{v_2^1, b_2^2\}$ for every supported bidder-2 bid. Applying the two maps defines

$$\widehat{\mathbf{b}}^1 \doteq (b_1^1, v_2^1 - 1), \quad \widehat{\mathbf{b}}^2 \doteq (b_1^2, b_2^2). \quad (28)$$

For bidder 1, the changed second coordinate in $\widehat{\mathbf{b}}^1$ remains below every supported bidder-2 first bid. For bidder 2, the second coordinate in $\widehat{\mathbf{b}}^2$ changes only when $b_1^2 = v_2^1$, in which case $\underline{b}_1^1 \geq v_2^1$ ensures that it remains losing. Thus Equation (28) maps each supported profile $\mathbf{b}$ to $\widehat{\mathbf{b}}$ without changing its allocation or payment.

Part 4: The case of $v_2^1 \geq v_2^2$. Part 3 constructed $\widehat{\mathbf{b}}$ for the first ordering of the bidders' second values. We now define the corresponding transformed profile for the remaining ordering and again

show that $\mathbf{b}$ and $\widehat{\mathbf{b}}$ have the same allocation, payments, and bidder utilities. The same argument applies after exchanging the bidder roles, though we have to be somewhat careful due to the asymmetric tie-breaking rule. Bidder 1's tie priority changes the cutoff by one: bidder 1 can preserve a baseline unit by tying bidder 2's largest feasible second bid $v_2^2 - 1$, whereas bidder 2 must bid at least v_2^2 to beat that bid. Hence $\underline{b}_1^1 = v_2^2 - 1$, $\underline{b}_1^2 \geq v_2^2$, and every supported bidder-1 bid satisfies $b_1^1 = \max\{v_2^2 - 1, b_2^1\}$. The maps define

$$\widehat{\mathbf{b}}^1 \doteq (b_1^1, b_1^1), \quad \widehat{\mathbf{b}}^2 \doteq (b_1^2, v_2^2 - 1). \quad (29)$$

Again, only coordinates that remain losing are changed, so $\mathbf{b}$ and $\widehat{\mathbf{b}}$ have the same allocations and payments. If $v_2^2 = 1$, the transformed bidder-2 second bid $\widehat{b}_2^2$ is simply zero.

Part 5: These maps preserve support optimality, CCE, and product structure. Parts 3 and 4 defined $\widehat{\mathbf{b}}$ for every supported profile $\mathbf{b}$ and showed that the transformation preserves its allocation, payment, bidder utilities, and welfare. We now pass from this profile map to the distribution $\widehat{\mathbf{p}}$ and verify that it preserves the solution-class properties used later. Let $\widehat{\mathbf{p}}$ be the distribution of $\widehat{\mathbf{b}}$ when $\mathbf{b} \sim \mathbf{p}$. Supported utilities and welfare are preserved at each profile as we showed above. Against any fixed deviation, the transformed opponent bid is weakly higher only in coordinates that remain losing on the transformed support. This can remove units from the deviator but cannot create a new win or reduce its payment in PAB. Since strict no overbidding makes every won unit weakly beneficial, no deviation becomes more profitable.

Every transformed supported action $\widehat{\mathbf{b}}^n$ therefore remains a best response, so $\widehat{\mathbf{p}} \in \mathcal{P}^*$. The same comparison preserves every CCE inequality. Finally, the maps act separately on the two bidder marginals, so they preserve product distributions and hence map MNE to MNE. $\square$

THEOREM 2 (PAB welfare hierarchy). *We have*

$$\text{PoA}_{\Delta(\widetilde{\mathcal{A}})} = 0 < 11/14 = \text{PoA}_{\mathcal{P}^*}, \quad \text{PoA}_{\text{CCE}} \leq \text{PoA}_{\text{CCE}_{CR}} < \text{PoA}_{\text{BR-CCE}} = \text{PoA}_{\text{MNE}} = 5/6.$$

Proof of Theorem 2. Proof roadmap. The proof has five parts. Part 1 applies Lemma 1 and reduces every potentially inefficient outcome to a comparison of the bidders' first-coordinate bids. Part 2 uses support optimality to determine their marginal bid distributions. Part 3 states the coupling and CCE inequalities and delegates the remaining exact integer optimization to the accompanying symbolic verification script. Part 4 proves tightness and the zero guarantee for correlated-rationalizable support alone. Part 5 gives a five-profile CCE_{CR} distribution below 5/6.

Part 1: Reduce supported bids to their first coordinates. Fix a valuation profile $\mathbf{v}$ and $\mathbf{p} \in \mathcal{P}^*(\mathbf{v})$. If every supported allocation is efficient, all positive lower bounds are immediate. Otherwise, apply Lemma 1 to obtain $\widehat{\mathbf{p}}$. Since $\widehat{\mathbf{p}}$ has the same welfare and the same relevant solution-class properties as $\mathbf{p}$, we continue to denote the transformed distribution by $\mathbf{p}$ to avoid carrying hats through the remaining calculation.

We first consider $v_2^1 < v_2^2$. The transformed bids are $(b^1, v_2^1 - 1)$ for bidder 1 and (b^2, b^2) for bidder 2. Bidder 2 always wins one baseline unit; the allocation splits when $b^1 \geq b^2$ and gives both units to bidder 2 when $b^1 < b^2$. Thus the entire welfare calculation will depend only on the probability of this last event.

Let ℓ_n be bidder n 's minimum supported first-coordinate bid. The reduction lemma gives $\ell_2 = v_2^1$ and $\ell_1 \geq \ell_2$. If $\ell_1 > \ell_2$, bidder 1 can optimally use ℓ_1 only if bidder 2 puts positive probability on ℓ_1 ; otherwise bidder 1 can lower to $\ell_1 - 1$ without changing its allocation. But bidder 2 can optimally use ℓ_1 only if bidder 1 puts positive probability on $\ell_1 - 1$; otherwise bidder 2 can lower to $\ell_1 - 1$ without changing whether it wins the second unit. This contradicts the definition of ℓ_1 . Hence $\ell_1 = \ell_2$; write their common value as $\underline{b}$ and let $q \doteq \Pr(b^1 < b^2 \mid \mathbf{b} \sim \mathbf{p})$. Every supported allocation is either split or gives both units to bidder 2, so

$$\text{WEL}(\mathbf{p}; \mathbf{v}) = v_1^1 + v_1^2 + (v_2^2 - v_1^1)q, \quad \text{OPT}(\mathbf{v}) = \max\{v_1^1 + v_1^2, v_1^2 + v_2^2\}. \quad (30)$$

Thus all welfare loss in this case is summarized by the single bundle probability q .

Part 2: Determine the marginal bid distributions from support optimality. Part 1 reduced welfare to the bundle probability q . We now determine which marginal distributions are compatible with support optimality; the later welfare bounds differ only in how much freedom they allow in coupling those marginals.

Let $\bar{b}$ be bidder 1's largest supported first-coordinate bid. The supported bid levels have no gaps. Indeed, if bidder 1 uses $b > \underline{b}$ but bidder 2 has no mass on b , bidder 1 can lower to $b - 1$ without changing its winning probability. If bidder 2 uses $b > \underline{b}$ but bidder 1 has no mass on $b - 1$, bidder 2 can lower to $b - 1$ without changing whether it wins the second unit. Repeating these implications gives

$$\text{Supp}(\mathbf{p}^1) = \{\underline{b}, \dots, \bar{b}\}, \quad \{\underline{b}, \dots, \bar{b}\} \subseteq \text{Supp}(\mathbf{p}^2) \subseteq \{\underline{b}, \dots, \bar{b} + 1\}.$$

Because all supported bids are best responses, they give the same expected utility. Bidder 2's minimum bid gives utility $v_1^2 - \underline{b}$, so comparison with any supported bid $b > \underline{b}$ gives

$$\Pr(b^1 \leq b - 1 \mid b^1 \sim \mathbf{p}^1) = \frac{b - \underline{b}}{v_2^2 - \underline{b}}, \quad \underline{b} < b \leq \bar{b}. \quad (31)$$

Let U_1^* denote bidder 1's best-response utility against $\mathbf{p}^2$. A supported bidder-1 bid b wins precisely when $b^2 \leq b$, so we have that

$$\Pr(b^2 \leq b \mid b^2 \sim \mathbf{p}^2) = \frac{U_1^*}{v_1^1 - b} \quad (\underline{b} \leq b \leq \bar{b}), \quad v_1^1 - \bar{b} - 1 \leq U_1^* \leq v_1^1 - \underline{b}. \quad (32)$$

The upper bound follows from the supported bid $\bar{b}$. The lower bound follows from the deviation $\bar{b} + 1$ whenever that bid is feasible and is otherwise trivial.

The endpoint of the support leaves only two integer cases. If $\bar{b} > \underline{b}$, Equation (31) at $b = \bar{b}$ and positive probability on the highest bid imply $v_2^2 - \underline{b} - 2(\bar{b} - \underline{b}) \geq 1$; when $\bar{b} = \underline{b}$, the same inequality follows from feasibility of bidder 2's minimum bid. Comparing bidder 2's minimum supported bid with the deviation $\bar{b} + 1$ gives the reverse bound of two whenever that deviation is feasible, while infeasibility gives an even smaller upper bound. Hence

$$v_2^2 - \underline{b} - 2(\bar{b} - \underline{b}) \in \{1, 2\}. \quad (33)$$

Thus every potentially inefficient support-optimal distribution has gap-free supported bid levels, marginals satisfying Equations (31) and (32), and one of the two cases in Equation (33). When $v_2^1 \geq v_2^2$, bidder 1 always wins one unit; exchanging bidder roles and shifting bidder 2's first-coordinate bid by one gives the same probability identities and the same integer optimization.

Part 3: Complete the welfare bounds by exact symbolic verification. Part 2 determined the marginal bid distributions and reduced the remaining possibilities to the two cases in Equation (33). It remains to bound how their coupling affects the bundle probability q and, when the coupling is also a CCE, how the two CCE deviations further restrict that probability. This part records those constraints and delegates only the remaining integer calculation to the accompanying verification script. For any coupling of the two marginals and any integer b , the event inclusions

$$\{b^1 \leq b\} \subseteq \{b^2 \leq b\} \cup \{b^1 < b^2\}, \quad \{b^2 \leq b\} \subseteq \{b^1 \leq b - 1\} \cup \{b^2 \leq b^1\} \quad (34)$$

bound the bundle probability q using only the marginal distributions in Equations (31) and (32). If the coupling is also a CCE, bidder 1's deviation to its highest supported bid and bidder 2's deviation to its lowest supported bid give the two additional inequalities needed when the split allocation and bidder 2's bundle, respectively, are efficient.

Substituting these constraints into Equation (30) leaves an exact integer optimization over the bidders' values and the smallest and largest supported bids. The accompanying script

`verify_welfare_bounds_symbolic.py` solves this optimization globally by symbolic case elimination. It has no finite search cutoff and uses no floating-point arithmetic or numerical tolerances: after branching over the two cases in Equation (33) and the two efficient allocations, it clears only denominators that are positive on the relevant branch and uses exact polynomial identities to certify the resulting inequalities. The role-reversed minimum-bid case is covered by the exact relabeling described at the end of Part 2. The script certifies

$$\frac{W_{EL}(\mathbf{p}; \mathbf{v})}{OPT(\mathbf{v})} \geq \frac{11}{14} \quad \text{for every } \mathbf{p} \in \mathcal{P}^*(\mathbf{v}), \quad \frac{W_{EL}(\mathbf{p}; \mathbf{v})}{OPT(\mathbf{v})} \geq \frac{5}{6} \quad \text{for every } \mathbf{p} \in \text{BR-CCE}(\mathbf{v}). \quad (35)$$

Consequently,

$$\text{PoA}_{\mathcal{P}^*}(2, 2) \geq \frac{11}{14}, \quad \text{PoA}_{\text{BR-CCE}}(2, 2) \geq \frac{5}{6}. \quad (36)$$

Every MNE is a BR-CCE, so the same 5/6 lower bound holds for MNE.

Part 4: Prove tightness and show that CR support alone gives no guarantee. The preceding parts established the universal lower bounds. The following finite constructions establish tightness and the zero guarantee for CR support alone. Their equilibrium and support constraints are verified by exact enumeration of every feasible bid vector in the accompanying script `verify_finite_examples.py`, using exact rational arithmetic.

For the tight 5/6 example, let $\mathbf{v}^1 = (2, 1)$ and $\mathbf{v}^2 = (3, 3)$, and consider $\mathbf{b}^1 = (1, 0)$ and $\mathbf{b}^2 = (1, 1)$. Exact enumeration verifies that this profile is a PNE. Its allocation is split, so welfare is 5 and $OPT = 6$. Therefore

$$\text{PoA}_{\text{BR-CCE}}(2, 2) = \text{PoA}_{\text{MNE}}(2, 2) = \frac{5}{6}.$$

For the tight 11/14 example, let $\mathbf{v}^1 = (4, 1)$ and $\mathbf{v}^2 = (7, 7)$. Use the marginals

$$\begin{aligned} p^1((1, 0)) &= \frac{1}{5}, & p^1((2, 0)) &= \frac{3}{10}, & p^1((3, 0)) &= \frac{1}{2}, \\ p^2((1, 1)) &= \frac{1}{3}, & p^2((2, 2)) &= \frac{1}{6}, & p^2((3, 3)) &= \frac{1}{2}. \end{aligned}$$

and the coupling

$$p((1, 0), (1, 1)) = \frac{1}{5}, \quad p((2, 0), (1, 1)) = \frac{2}{15}, \quad p((2, 0), (2, 2)) = \frac{1}{6}, \quad p((3, 0), (3, 3)) = \frac{1}{2}.$$

Exact enumeration verifies that every marginally supported bid is a best response to the opponent marginal, so $\mathbf{p} \in \mathcal{P}^*$. Every supported profile splits the units, and hence

$$\frac{W_{EL}(\mathbf{p}; \mathbf{v})}{OPT(\mathbf{v})} = \frac{4 + 7}{7 + 7} = \frac{11}{14}.$$

Together with Equation (36), this proves $\text{PoA}_{\mathcal{P}^*}(2, 2) = 11/14$.

Finally, let $\mathbf{v}^1 = (H, 1)$ and $\mathbf{v}^2 = (4, 4)$, and put probability one on $((1, 0), (2, 2))$. Exact best-response enumeration verifies that $(1, 0), (2, 0)$ for bidder 1 and $(1, 1), (2, 2)$ for bidder 2 form a set satisfying the CR best-response condition; the verification output records, for each listed action, an opponent action against which it is a best response. Bidder 2 wins both units, so

$$\frac{W_{\text{EL}}(\mathbf{p}; \mathbf{v})}{\text{OPT}(\mathbf{v})} = \frac{8}{H+4} \rightarrow 0 \implies \text{PoA}_{\Delta(\tilde{\mathcal{A}})}(2, 2) = 0.$$

Part 5: Show that CCE_{CR} can still be worse than BR-CCE. Part 4 showed that CR support alone can have vanishing welfare. The final construction shows that the gap remains after imposing CCE. Let bidder 1 have tie priority, let $\mathbf{v}^1 = (6, 6)$ and $\mathbf{v}^2 = (8, 8)$. Exact enumeration over all feasible deviations verifies the CCE constraints, the CR constraints, and the failure of support optimality; hence $\mathbf{p} \in \text{CCE}_{CR} \setminus \text{BR-CCE}$. The two tables below record the distribution and the three coordinate-payoff expressions evaluated by the verifier.

$\mathbf{b}^1$	$\mathbf{b}^2$	$p(\mathbf{b}^1, \mathbf{b}^2)$								
(0, 0)	(1, 1)	1/7								
(0, 0)	(2, 0)	1/7								
(2, 2)	(3, 3)	4/35								
(4, 4)	(4, 4)	6/35								
(5, 5)	(5, 5)	3/7								

b	0	1	2	3	4	5	6	7
$(6-b) \Pr(b_2^2 \leq b \mid \mathbf{b}^2 \sim \mathbf{p}^2)$	6/7	10/7	8/7	6/5	8/7	1	-	-
$(6-b) \Pr(b_1^2 \leq b \mid \mathbf{b}^2 \sim \mathbf{p}^2)$	0	5/7	8/7	6/5	8/7	1	-	-
$(8-b) \Pr(b_1^1 < b \mid \mathbf{b}^1 \sim \mathbf{p}^1)$	0	2	12/7	2	8/5	12/7	2	1

The welfare at the five profiles is 16, 14, 16, 12, and 12, respectively, so $\frac{W_{\text{EL}}(\mathbf{p}; \mathbf{v})}{\text{OPT}(\mathbf{v})} = \frac{233}{280} < \frac{5}{6}$ and consequently, we have the stated PoA guarantees.

$$\text{PoA}_{\text{CCE}}(2, 2) \leq \text{PoA}_{\text{CCE}_{CR}}(2, 2) \leq \frac{233}{280} < \text{PoA}_{\text{BR-CCE}}(2, 2) = \text{PoA}_{\text{MNE}}(2, 2) = \frac{5}{6}.$$

Combining Parts 3–5 proves every equality and inequality in Theorem 2. $\square$

C. Small-Gap Exploration Proofs

We now prove the necessary stay-frequency condition and the forgiving non-convergence theorem.

C.1. Proof of Theorem 3

THEOREM 3 (Small Gap Exploration: Necessary Stay Frequency). *Let $\mathbf{p}$ be a (possibly correlated) distribution and $a \in S^n(\mathbf{p})$ be a substitute for player n . If GBA dynamics converge to $\mathbf{p}$, then*

$$P_t^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \gg t^{\alpha_n}.$$

Proof. Suppose, toward contradiction, that the dynamics converge to $\mathbf{p}$ but the stay frequency does not eventually exceed the tolerance. Then there are a constant $C < \infty$ and horizons $t_k \rightarrow \infty$ such that

$$P_{t_k}^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \leq C t_k^{\alpha_n}. \quad (37)$$

We first show two claims. The first says that in almost every round, *some* supported action of player n is still played with non-vanishing probability. This action may change from round to round, so current strategies can cycle even though their time average converges, as can happen when the limit is not an NE. The second says that in these rounds, at least in the second half of each horizon t_k , the substitute's dynamic gap is within tolerance. Small-gap exploration then gives the substitute positive empirical frequency, contradicting its absence from the limiting support.

Claim 1. On all but $o(t)$ of the first t rounds, there exists an action $b_\tau \in \text{Supp}(\mathbf{p}^n)$ such that

$$p_\tau^n(b_\tau) \geq \frac{1}{2|\text{Supp}(\mathbf{p}^n)|}. \quad (38)$$

Proof of Claim 1. To prove this, we note that because the dynamics converge to $\mathbf{p}$, every action outside of the support of $\mathbf{p}^n$ —including substitutes such as a —must have $o(t)$ cumulative frequency:

$$P_t^n(a) = \sum_{\tau < t} p_\tau^n(a) = \sum_{\tau < t} \left(1 - \sum_{b \neq a} p_\tau^n(b) \right) \leq \sum_{\tau < t} \left(1 - \sum_{b \in \text{Supp}(\mathbf{p}^n)} p_\tau^n(b) \right) = o(t), \quad (39)$$

Every round τ on which $\sum_{b \in \text{Supp}(\mathbf{p}^n)} p_\tau^n(b) < \frac{1}{2}$ makes the corresponding summand larger than $1/2$ in Equation (39). There can therefore be only $o(t)$ such rounds. On every remaining round, the finite support contains some action b_τ with $p_\tau^n(b_\tau) \geq \frac{1}{2|\text{Supp}(\mathbf{p}^n)|}$. $\square$

Claim 2. There is a constant $\zeta < \infty$ such that, for every sufficiently large k and every round $\tau \in [t_k/2, t_k)$, if some $b \in \text{Supp}(\mathbf{p}^n)$ satisfies $p_\tau^n(b) \geq 1/(2|\text{Supp}(\mathbf{p}^n)|)$, then

$$\Delta_\tau^n(a) \leq \zeta \tau^{\alpha_n}. \quad (40)$$

Proof of Claim 2. First, large-gap pruning gives constants $K < \infty$ and $T_1 < \infty$ such that, for every $\tau \geq T_1$ and every $b \in \text{Supp}(\mathbf{p}^n)$,

$$p_\tau^n(b) \geq \frac{1}{2|\text{Supp}(\mathbf{p}^n)|} \implies \Delta_\tau^n(b) \leq K \tau^{\alpha_n}. \quad (41)$$

Indeed, otherwise some fixed supported action (the support is finite) would keep probability at least $\frac{1}{2|\text{Supp}(\mathbf{p}^n)|}$ on infinitely many rounds while its gap grows faster than the tolerance along those rounds. Large-gap pruning and Equation (7) would then force its probability to zero along those rounds, a contradiction. (To apply large-gap pruning, which is stated for full sequences, set $\Delta_t = t^{(1+\alpha_n)/2}$ on all other rounds.) We next compare the gap of supported action b with the gap of the substitute a .

For every $b \in \text{Supp}(\mathbf{p}^n)$, we can rewrite this gap as the weighted sum of the gap between b and a at each round $s < \tau$:

$$\begin{aligned}
 U_\tau^n(b) - U_\tau^n(a) &= U_0^n(b) - U_0^n(a) + \sum_{s < \tau} \sum_{\mathbf{a}^{-n}} (u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})) \prod_{i \neq n} p_s^i(a^i) \\
 &= O(1) + \sum_{s < \tau} \sum_{\mathbf{a}^{-n}} (u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})) \prod_{i \neq n} p_s^i(a^i) \\
 &= O(1) + \sum_{s < \tau} \sum_{\mathbf{a}^{-n} \in \text{STAY}^{-n}(a; \mathbf{p})} (u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})) \prod_{i \neq n} p_s^i(a^i) \\
 &\quad + \sum_{s < \tau} \sum_{\mathbf{a}^{-n} \notin \text{STAY}^{-n}(a; \mathbf{p})} (u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})) \prod_{i \neq n} p_s^i(a^i) \tag{42} \\
 &\leq O(1) + \sum_{s < \tau} \sum_{\mathbf{a}^{-n} \in \text{STAY}^{-n}(a; \mathbf{p})} (u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})) \prod_{i \neq n} p_s^i(a^i) \\
 &\leq O(1) + c_0 \sum_{s < \tau} \sum_{\mathbf{a}^{-n} \in \text{STAY}^{-n}(a; \mathbf{p})} \prod_{i \neq n} p_s^i(a^i) \\
 &= O(1) + c_0 P_\tau^{-n}(\text{STAY}^{-n}(a; \mathbf{p})),
 \end{aligned}$$

where $c_0 \doteq \max_{\mathbf{a}^{-n}, b \in \text{Supp}(\mathbf{p}^n)} [u^n(b, \mathbf{a}^{-n}) - u^n(a, \mathbf{a}^{-n})]^+ < \infty$. The initialization differences are bounded uniformly over the finite support. The payoff difference is nonpositive outside $\text{STAY}^{-n}(a; \mathbf{p})$, which justifies discarding those terms in the upper bound.

Cumulative stay frequency is nondecreasing. Thus, for every $\tau \in [\lceil t_k/2 \rceil, t_k)$, Equation (37) gives

$$P_\tau^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \leq P_{t_k}^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \leq C t_k^{\alpha_n} \leq 2^{\alpha_n} C \tau^{\alpha_n}.$$

Absorbing the finite initialization term into τ^{α_n} , Equation (42) therefore implies, for every sufficiently large k and every supported b ,

$$U_\tau^n(b) - U_\tau^n(a) \leq (1 + 2^{\alpha_n} c_0 C) \tau^{\alpha_n} \quad (\lceil t_k/2 \rceil \leq \tau < t_k). \tag{43}$$

On any such round satisfying the probability premise of Claim 2, Equation (41) yields

$$\Delta_\tau^n(a) = \Delta_\tau^n(b) + U_\tau^n(b) - U_\tau^n(a) \leq (K + 1 + 2^{\alpha_n} c_0 C) \tau^{\alpha_n}. \tag{44}$$

This proves Claim 2 with $\zeta = K + 1 + 2^{\alpha_n} c_0 C$. $\square$

We now combine the two claims on $[\lceil t_k/2 \rceil, t_k)$. By Claim 1, at least $t_k/2 - o(t_k) - O(1)$ rounds in this interval have a supported action with the stated probability, and by Claim 2, the substitute's gap is at most $\zeta \tau^{\alpha_n}$ on each of them once k is large. Since this gap is within tolerance, small-gap

exploration (with monotonicity of γ_τ^n and Equation (7)) gives a constant $\eta > 0$, the same for every k , with $p_\tau^n(a) \geq \eta$ on those rounds. Hence

$$P_{t_k}^n(a) \geq \eta(t_k/2 - o(t_k) - O(1)) = \Omega(t_k),$$

contradicting Equation (39). Therefore

$$\frac{P_t^{-n}(\text{STAY}^{-n}(a; \mathbf{p}))}{t^{\alpha_n}} \longrightarrow \infty. \quad \square$$

C.2. Proof of Theorem 4

THEOREM 4 (Non-Convergence of Forgiving Dynamics to Certain Limits). *Let $\mathbf{p}$ be a distribution and let $a \in S^n(\mathbf{p})$. Suppose that every profile in the stay set from Definition 7 contains an opponent action with positive static gap:*

$$\mathbf{a}^{-n} \in \text{STAY}^{-n}(a; \mathbf{p}) \implies \delta^i(a^i; \mathbf{p}) > 0 \text{ for some } i \neq n$$

Then n -forgiving GBA dynamics cannot converge to $\mathbf{p}$.

Proof. We proceed by contradiction. Suppose that the dynamics converge to $\mathbf{p}$. Fix $\mathbf{a}^{-n} \in \text{STAY}^{-n}(a; \mathbf{p})$, with the stay set defined in Definition 7. By the premise of Theorem 4, this profile contains an opponent action a^i with $\delta^i(a^i; \mathbf{p}) > 0$. Proposition 12 gives constants $d > 0$ and $T < \infty$ such that $\Delta_\tau^i(a^i) \geq d\tau$ for every $\tau \geq T$. Using monotonicity of γ_τ^i and n -forgiveness from Definition 8,

$$P_t^{-n}(\mathbf{a}^{-n}) \leq P_t^i(a^i) \lesssim 1 + \sum_{\tau < t} \gamma_\tau^i(\Delta_\tau^i(a^i)) \lesssim 1 + \sum_{\tau < t} \gamma_\tau^i(d\tau) \lesssim t^{\alpha_n}. \quad (45)$$

There are only finitely many profiles in the stay set, so Equation (45) gives $P_t^{-n}(\text{STAY}^{-n}(a; \mathbf{p})) \lesssim t^{\alpha_n}$. The stay frequency therefore stays within tolerance, contradicting Theorem 3. $\square$

C.3. Limitations of Forgiving Dynamics

Theorem 3 rules out convergence whenever the stay frequency stays within tolerance, and Theorem 4 gives a sufficient condition for this: forgiving dynamics, together with a positive-gap opponent action in every stay profile. Neither ingredient always holds. Not every GBA is forgiving (Proposition 13), and not every weak equilibrium has a positive-gap opponent action in every stay profile (Proposition 14).

PROPOSITION 13 (Forced exploration may violate forgiveness). *There exist ϵ_t -buffered GBAs that are n -forgiving for exactly one player n , and there exist ϵ_t -buffered GBAs that are not forgiving for any player.*

Proof. Use two-player buffered polynomial weights with $\alpha_1 = \alpha_2 = 1/2$ and $\epsilon_t^i = t^{-s_i}$. For a linear-gap action, the rate calculation in Proposition 9 gives

$$\sum_{\tau < t} \gamma_{\tau}^i(\kappa\tau) \asymp t^{1-m_i}, \quad m_i = \min\{\beta_i/2, s_i\}.$$

Thus player n is forgiving exactly when $m_i \geq 1/2$ for the other player $i \neq n$.

Take $\beta_1 = \beta_2 = 4$, $s_1 = .40$, and $s_2 = .95$. Then $m_1 = .40$ and $m_2 = .95$, so player 1 is forgiving while player 2 is not. Next take $\beta_1 = \beta_2 = 6$ and $s_1 = s_2 = .45$. Then $m_1 = m_2 = .45$, so neither player is forgiving. The exploration floor vanishes in both examples, so the rules remain GBAs. $\square$

PROPOSITION 14 (A weak equilibrium outside the forgiving condition). *There exists a weak NE with substitutes that does not satisfy the forgiving static condition.*

Proof. Consider the symmetric two-player game with actions A, B and payoffs

$$u(A, A) = (0, 0), \quad u(A, B) = (2, 0), \quad u(B, A) = (0, 2), \quad u(B, B) = (1, 1).$$

The profile (A, A) is a weak PNE under Definitions 1 and 2. Against opponent action A , both A and B give payoff zero, so B is a substitute for each player and neither action has positive static gap. For player 1, Definition 7 gives $\text{STAY}^{-1}(B; (A, A)) = \{B\}$ because $u^1(A, B) = 2 > 1 = u^1(B, B)$. The only stay action of player 2 therefore has zero static gap at (A, A) , so the forgiving static condition fails. Interchanging the two players shows that player 2 also has a zero-static-gap stay action, so the premise of Theorem 4 fails for both players. $\square$

D. Warm-Up Application Proofs and Simulations

This section pertains to the bilateral-trade example in Section 5.1 and the FPA example in Section 5.3. It supplies the remaining proof details and simulation descriptions so that those sections can focus on the small-gap results.

D.1. Warm-Up Applications

This subsection pertains to the bilateral-trade proposition in Section 5.1. Proposition 4 in Section 5.1 explains why GBA dynamics avoid no-trade equilibria; here we give the contrasting MBA and no-regret constructions. We provide the missing proofs regarding our warm-up applications.

D.1.1. Mean-based and no-regret convergence to no trade

PROPOSITION 5 (MBAs and no-regret dynamics can converge to no trade). *There exist MBA dynamics and no-regret dynamics whose play converges to the endpoint no-trade equilibrium (c, v) in the bilateral-exchange game.*

Proof.

Mean-based dynamics. Fix any $\beta_t \rightarrow 0$. Consider a deterministic tie-breaking rule satisfying the MBA condition in Definition 5: it assigns probability at most β_t to every action whose gap exceeds $t\beta_t$ and assigns all remaining probability to a fixed tie-breaking action. For the buyer, choose tie-breaking action c ; for the seller, choose tie-breaking action v . If play starts at (c, v) and both agents continue to play these actions, then every feasible buyer bid earns zero against ask v , and every feasible seller ask earns zero against bid c . Hence all gaps remain zero, the MBA rule is never forced to prune the tie-breaking action, and play stays at (c, v) forever.

No-regret dynamics. Use PGD on the simplex with any standard no-regret step-size sequence, initialized at the pure endpoint action: the buyer at bid c and the seller at ask v . Along the path (c, v) , every feasible buyer bid earns zero against ask v , and every feasible seller ask earns zero against bid c . Hence the payoff-gradient vector is constant across actions for each player, and Equations (11) and (14) show that the projection leaves the current vertex unchanged. Regret is therefore identically zero. Thus no-regret dynamics can also converge to the endpoint no-trade equilibrium. $\square$

D.2. Proof of Proposition 6

We prove that the weak FPA equilibrium can persist under weaker dynamics even though forgiving GBAs rule it out.

PROPOSITION 6 (Weaker dynamics can converge to the weak FPA equilibrium). *In the two-bidder FPA, there exist MBA dynamics, no-regret dynamics, and non-forgiving GBA dynamics whose play converges to the weak PNE $b^1 = b^2 = v - 2$.*

Proof. Consider the two-top-bidder FPA at the weak PNE $b^1 = b^2 = v - 2$. Against an opponent bid of $v - 2$, bids $v - 2$ and $v - 1$ both give payoff 1, while every lower bid gives payoff 0.

FTL / mean-based dynamics. Give each bidder a finite initial prior $U_0^n(v - 2) > U_0^n(v - 1)$ and $U_0^n(v - 2) > U_0^n(b)$ for every $b \leq v - 3$. If both bidders play $v - 2$, then $v - 2$ and $v - 1$ earn the same payoff at every subsequent round, and all lower bids earn strictly less. Hence the initial strict lead of $v - 2$ over $v - 1$ is never erased, and $v - 2$ remains the unique leader forever. FTL is an MBA, so this gives an MBA path converging to the weak PNE.

PGD / no-regret dynamics. Initialize both bidders at the point mass on $v - 2$. Suppose the opponent remains at $v - 2$. The payoff vector has value 1 at $v - 2$ and $v - 1$, and value 0 at every lower bid. The PGD vector before projection is therefore

$$\mathbf{1}_{v-2} + \eta_t(\mathbf{1}_{v-2} + \mathbf{1}_{v-1}),$$

up to the zero coordinates on lower bids. Projecting this vector onto the simplex subtracts η_t from the two positive candidate coordinates and truncates the lower coordinates at zero, leaving exactly $\mathbf{1}_{v-2}$. By induction, PGD remains at $v - 2$ forever. This is a standard no-regret full-information dynamic, but it is not a GBA.

A non-forgiving GBA. Fix $\alpha \in (0, 1)$ and $0 < s < 1 - \alpha$. Both bidders run Hedge with a small amount of uniform exploration:

$$p_t^n(b) = (1 - \epsilon_t) \frac{e^{U_t^n(b)/t^\alpha}}{\sum_{a=0}^{v-1} e^{U_t^n(a)/t^\alpha}} + \frac{\epsilon_t}{v}, \quad \epsilon_t = \frac{t^{-s}}{4v(v+1)}.$$

This is a GBA with tolerance exponent α and sensitivity $\gamma_t(\Delta) = (1 - \epsilon_t)e^{-\Delta/t^\alpha} + \epsilon_t$, since $\gamma_t(\Delta_t^n(b))/v \leq p_t^n(b) \leq \gamma_t(\Delta_t^n(b))$. It is not forgiving: exploration alone gives every action, including linear-gap actions, cumulative frequency of order t^{1-s} , which exceeds the tolerance t^α . That is, for every $d > 0$,

$$\sum_{\tau < t} \gamma_\tau(d\tau) \asymp t^{1-s} \gg t^\alpha.$$

Both bidders start from the same prior, $U_0^n(v-2) = L$ and $U_0^n(b) = 0$ for $b \neq v-2$, so by symmetry they always play the same strategy $\mathbf{p}_t$, and we drop player superscripts. The idea is that exploration works in favor of the weak equilibrium. Against the equilibrium bid $v - 2$, bidding $v - 2$ and the substitute $v - 1$ earn the same. When the opponent explores a bid below $v - 2$, bidding $v - 2$ earns one more than $v - 1$; when the opponent explores $v - 1$, the substitute earns $1/2$ more. There are $v - 2$ low bids and only one high bid, so $v - 2$ pulls ahead of $v - 1$ at a rate proportional to ϵ_t . Its lead grows like t^{1-s} , faster than the tolerance t^α , so the substitute is pruned. The prior L covers the early rounds.

To make this precise, choose L large enough that

$$G_t \doteq L + \frac{1}{12} \sum_{1 \leq \tau < t} \epsilon_\tau \geq t^\alpha \log(6/\epsilon_t) \quad (t \geq 1), \quad (46)$$

which is possible because $1 - s > \alpha$. We show that $v - 2$ always leads every other bid by at least G_t , i.e., $U_t(v-2) - U_t(b) \geq G_t$ for all $b \neq v-2$. The choice of L gives this at $t = 1$. Suppose it holds at

round t . Then the Hedge part of the update gives each bid $b \neq v - 2$ weight at most $e^{-G_t/t^\alpha} \leq \epsilon_t/6$, so

$$p_t(b) \leq \frac{\epsilon_t}{6} + \frac{\epsilon_t}{v} \leq \epsilon_t, \quad 1 - p_t(v - 2) \leq v\epsilon_t \leq \frac{1}{4(v + 1)}.$$

Almost all probability is on $v - 2$, and the other bids are played only through exploration. Using $p_t(b) \geq \epsilon_t/v$ for the low bids, $v - 2$ beats the substitute by

$$\begin{aligned} u(v - 2, \mathbf{p}_t) - u(v - 1, \mathbf{p}_t) &= \sum_{b=0}^{v-3} p_t(b) - \frac{1}{2}p_t(v - 1) \\ &\geq \left(\frac{v - 2}{v} - \frac{1}{2v} - \frac{1}{12} \right) \epsilon_t \geq \frac{\epsilon_t}{12}. \end{aligned}$$

It also beats every lower bid $b \leq v - 3$: bidding $v - 2$ earns 1 more when the opponent bids $v - 2$, and loses at most v otherwise, so

$$u(v - 2, \mathbf{p}_t) - u(b, \mathbf{p}_t) \geq 1 - (v + 1)(1 - p_t(v - 2)) \geq \frac{3}{4} \geq \frac{\epsilon_t}{12}.$$

So the lead of $v - 2$ over every other bid grows by at least $\epsilon_t/12$ in round t , which gives the claim at round $t + 1$. Finally, $1 - p_t(v - 2) \leq v\epsilon_t \rightarrow 0$, so both bidders' play converges to the weak PNE $b^1 = b^2 = v - 2$. $\square$

D.3. Warm-up simulations

Section 5.1 and Section 5.3 use bilateral trade and the two-top-bidder FPA as transparent examples of the small-gap results. We include the simulation details here to isolate the multi-unit experiments in Section 6.

We initialize each bilateral-trade simulation at the endpoint no-trade equilibrium (c, v) , and we initialize each two-top-bidder FPA simulation at the weak PNE $v - 2$ using the finite prior $U_0^n(v - 2) = 10$ and $U_0^n(a) = 0$ for all other bids. Hedge is included as the canonical no-regret GBA and is run with learning rate $\eta_t = t^{-1/2}$. Polynomial weights is included to show that the behavior is not merely a consequence of no-regret, since polynomial weights is a GBA but is not generally no-regret; we use $\alpha = .25$ and $\beta = 2$, which satisfies the forgiving condition. The buffered rows test the role of forgiveness. Buffered - Forgiving uses the buffered polynomial rule with $\alpha = .5$, $\beta = 4$, and $s = .80$, while Buffered - Not forgiving uses $\alpha = .5$, $\beta = 6$, and $s = .45$. Both buffered rows use polynomial weights rather than the buffered Hedge rule from Proposition 6. Finally, FTL and PGD are included as non-GBA comparison dynamics. PGD uses step size $\eta_t = .25t^{-1/2}$; since PGD never leaves its starting point here, the step size does not matter.

Table 5 Warm-up dynamics. Bilateral trade uses 50 random instances and $T = 5000$ rounds. The FPA uses $v \in \{3, \dots, 15\}$ and $T = 10^5$ expected-feedback rounds. Each entry is the outcome reached in every run by the final round; FPA entries give the equilibrium bid that play ends up near.

Algorithm	Bilateral trade	Two-top-bidder FPA
Hedge/Poly Weights	Always Trade	$v - 1$
Buffered - Forgiving	Always Trade	$v - 1$
Buffered - Not forgiving	Always Trade	$v - 2$
FTL / Projected GD	No Trade	$v - 2$

The bilateral-trade column illustrates Proposition 4: all GBA-style dynamics in the table reach positive trade from the no-trade initialization, while FTL and projected GD remain at No Trade. In the FPA, Hedge, polynomial weights, and the forgiving buffered rule all move to the high-bid equilibrium $v - 1$, consistent with Corollary 2. The non-forgiving buffered rule stays near the weak equilibrium $v - 2$, and FTL and PGD stay exactly at $v - 2$. Proposition 6 proves this kind of convergence for buffered Hedge; the simulation shows the same behavior for buffered polynomial weights. Thus the table separates the effect of small-gap exploration from both no-regret learning and non-GBA comparison dynamics.

E. Multi-Unit Auction Details

This appendix supports Section 6 with additional experiments and the full UPA proof.

E.1. Additional capped threshold-bid experiments

This subsection contains additional simulations under a threshold-bid restriction:

$$\mathbf{b}^n(b) \doteq (\min\{b, \max\{0, \lfloor v_m^n \rfloor - 1\}\})_{m=1}^M, \quad b \in \mathbb{Z}_{\geq 0}.$$

This interface is especially natural for PAB because [32] find that no-regret learning dynamics in repeated PAB auctions typically move toward this thresholded bid-curve class. The right reduced interface is less clear for UPA because losing bids can affect prices. The smaller action space lets us run larger instances with $M = 10$, $V = 20$, and $N \in \{2, 3\}$. The direction of the results is similar to the experiments in Table 3, but the differences are smaller. We note that these capped threshold-bid experiments use a restricted action space and thus are not covered by Theorem 5.

Table 6 Capped threshold-bid experiments for $V = 20$ over 25 random instances, with $M = 10$. Entries are mean/worst normalized by OPT. For $1 - \text{WEL}$, worst means largest welfare loss; for $\text{REV}_{\max}$, worst means largest maximum revenue; and for $\text{REV}_{\min}$, worst means smallest minimum revenue. Static benchmark rows are shaded in gray. For Hedge, the same time-average revenue is summarized under both revenue columns; its $\text{REV}_{\max}$ worst entry is the largest realized revenue across instances, while its $\text{REV}_{\min}$ worst entry is the smallest realized revenue.

	$N = 2, V = 20$			$N = 3, V = 20$		
	$1 - \text{WEL}$	$\text{REV}_{\max}$	$\text{REV}_{\min}$	$1 - \text{WEL}$	$\text{REV}_{\max}$	$\text{REV}_{\min}$
PAB						
CCE	0.095 / 0.147	0.432 / 0.636	0.251 / 0.179	0.179 / 0.198	0.727 / 0.823	0.315 / 0.224
CCE _{CR}	0.042 / 0.100	0.408 / 0.595	0.365 / 0.180	0.010 / 0.051	0.711 / 0.796	0.696 / 0.493
BR-CCE	0.036 / 0.073	0.399 / 0.565	0.392 / 0.194	0.007 / 0.032	0.711 / 0.796	0.704 / 0.500
Hedge	0.032 / 0.069	0.396 / 0.565	0.396 / 0.195	0.006 / 0.029	0.707 / 0.793	0.707 / 0.500
UPA						
CCE	0.050 / 0.106	0.334 / 0.548	0.216 / 0.000	0.016 / 0.076	0.684 / 0.795	0.614 / 0.330
CCE _{CR}	0.044 / 0.106	0.334 / 0.548	0.226 / 0.000	0.009 / 0.063	0.684 / 0.795	0.654 / 0.343
BR-CCE	0.040 / 0.106	0.300 / 0.544	0.226 / 0.000	0.009 / 0.063	0.683 / 0.795	0.654 / 0.343
Hedge	0.022 / 0.077	0.293 / 0.544	0.293 / 0.000	0.002 / 0.020	0.674 / 0.795	0.674 / 0.473

E.2. UPA simulations

In Figure 1, we largely borrow the experimental setup of [32]. We use Hedge dynamics with tolerance $t^{1/2}$, expected feedback, uniform initialization, 3 bidders, 10 items, and integer values in $\{1, \dots, 10\}$. Curves and bands show the 10th, 25th, 50th, 75th, and 90th percentiles over 100 trials of time-averaged revenue and welfare.

E.3. UPA proof details

We prove the UPA theorem in Section 6.3. The proof first identifies price-setting losing bids and then verifies the forgiving condition for the resulting price-lowering substitute. We begin with an allocation-preservation lemma.

LEMMA 2 (Losing-bid lowering preserves allocations). *Fix any UPA bid profile $\mathbf{b}$ with deterministic tie-breaking, and let $W(\mathbf{b})$ be the set of the M winning bids at $\mathbf{b}$. Suppose $\mathbf{c}$ is obtained from $\mathbf{b}$ by leaving every bid in $W(\mathbf{b})$ unchanged and weakly lowering only bids outside $W(\mathbf{b})$. Then $W(\mathbf{c}) = W(\mathbf{b})$. In particular, every bidder's allocation is unchanged, and the uniform price can only weakly decrease.*

Proof. Every winning bid at $\mathbf{b}$ is left unchanged. Every modified bid was losing at $\mathbf{b}$ and is weakly lowered, so it cannot overtake any bid that was winning at $\mathbf{b}$. Therefore the same M

bids remain the top M bids after the modification, and thus, allocations are unchanged. Since the $(M + 1)$ -st bid is selected from a set in which the old winning bids are unchanged and the old losing bids are weakly lower, the clearing price weakly decreases, and all bidders' utilities weakly increase. $\square$

THEOREM 5 (Positive-revenue UPA PNE cannot be limits). *Consider UPA with integer bids and no overbidding such that:*

1. *Each bidder's valuations are generic: no bidder is indifferent between any two bids with different allocations—e.g., valuations are drawn from an absolutely continuous density.*
2. *Each bidder is non-spectating: bidder n guarantees allocation by bidding $\mathbf{b}^n = \lceil \mathbf{v}^n \rceil - \mathbf{1}$, its highest feasible bid (here $\mathbf{1}$ is the all-ones vector).*

Then, forgiving GBA dynamics cannot converge to any positive-revenue PNE.

Proof. At a high level, the proof has 5 parts. First, using our non-spectating assumption, we show that any PNE $\mathbf{b}$ must have positive allocation—denoted by $m_i > 0$ —for each bidder. Second, we argue that for positive revenue $\mathbf{b}$, the above positive allocation guarantee means that at least two different bidders have losing bids at exactly the clearing price $\pi > 0$. Thirdly, we design a substitute bid $\mathbf{b}_*^i$ using the fact that these price setting bidders can weakly lower their losing bids without either sacrificing allocation or increasing their cost. By our genericity assumption, we rule out the possibility of price *increasing* substitutes, where bidders increase their losing bids in order to improve allocation just enough to offset the increased costs. Using this fact, we show that every stay profile must contain a dominated action, thereby guaranteeing non-convergence of forgiving dynamics.

Step 1: every bidder has positive allocation. Fix a positive-revenue PNE $\mathbf{b}$. Let $\pi > 0$ be its clearing price, equal to the highest rejected bid, and let m_i be bidder i 's allocation at $\mathbf{b}$. If bidder i received no allocation under $\mathbf{b}$, they could deviate to $\mathbf{b}^i = \lceil \mathbf{v}^i \rceil - \mathbf{1}$ and guarantee positive allocation and utility. This contradicts that $\mathbf{b}$ is a PNE. Therefore $m_i > 0$ for all i .

Step 2: there are at least two price-setting losing bidders. Define the set of price-setting losing bidders by

$$\mathcal{N}_\pi \doteq \{i : \text{bidder } i \text{ has a losing bid equal to } \pi \text{ at } \mathbf{b}\}. \quad (47)$$

Suppose for contradiction that $\mathcal{N}_\pi = \{i\}$ contains only one bidder. If bidder i lowers every losing bid equal to π to at most $\pi - 1$ and leaves all winning bids fixed, its allocation is unchanged. Since

bidder i was the only bidder with a losing bid equal to π , the clearing price strictly decreases. Since $m_i > 0$, this strictly increases bidder i 's utility, contradicting the PNE property. Thus, $|\mathcal{N}_\pi| \geq 2$.

Step 3: price-lowering bids are substitutes. For each $i \in \mathcal{N}_\pi$, define $\mathbf{b}_*^i$ by keeping all winning bids of bidder i fixed and replacing every losing bid equal to π by $\pi - 1$; all other losing bids are left unchanged. The bid vector remains feasible and nonincreasing. Since $|\mathcal{N}_\pi| \geq 2$, when bidder i alone switches from $\mathbf{b}^i$ to $\mathbf{b}_*^i$ while opponents play $\mathbf{b}^{-i}$, at least one opponent still has a losing bid equal to π . Lemma 2 preserves bidder i 's allocation, and the clearing price remains π . Therefore

$$u^i(\mathbf{b}_*^i, \mathbf{b}^{-i}) = u^i(\mathbf{b}^i, \mathbf{b}^{-i}), \quad (48)$$

so $\mathbf{b}_*^i \in \mathcal{S}^i(\mathbf{b})$ by Definition 2. Fix a bidder $n \in \mathcal{N}_\pi$ and use $\mathbf{b}_*^n$ as bidder n 's substitute.

Step 3a: lowering losing bids cannot change allocations. We will repeatedly apply Lemma 2. In particular, whenever a set of price-setting losing bidders replace their losing π -bids by $\pi - 1$ while keeping winning bids fixed, allocations are unchanged and the only possible payment effect is to weakly lower the uniform price.

Step 3b: zero-gap opponent actions preserve that opponent's allocation. Let $i \neq n$ and suppose $\delta^i(\mathbf{a}^i; \mathbf{b}) = 0$. Since $\mathbf{b}$ is a PNE, both $\mathbf{a}^i$ and $\mathbf{b}^i$ are best responses to $\mathbf{b}^{-i}$. By genericity, the profiles $(\mathbf{a}^i, \mathbf{b}^{-i})$ and $(\mathbf{b}^i, \mathbf{b}^{-i})$ must give bidder i the same allocation, and thus also the same clearing price. In particular, bidder i still wins exactly m_i units when it alone switches to $\mathbf{a}^i$.

Step 4: every stay profile contains a positive-static-gap opponent action. We claim that

$$\mathbf{a}^{-n} \in \text{STAY}^{-n}(\mathbf{b}_*^n; \mathbf{b}) \implies \exists i \neq n \text{ such that } \delta^i(\mathbf{a}^i; \mathbf{b}) > 0. \quad (49)$$

Suppose instead that every opponent component has zero static gap. Fix one of bidder n 's losing bids equal to π that is lowered in $\mathbf{b}_*^n$. For any opponent $i \neq n$, consider the unilateral profile $(\mathbf{a}^i, \mathbf{b}^{-i})$. Step 3b of this proof implies that bidder i wins exactly m_i units. Since all bids of the other players are fixed, the remaining $M - m_i$ winners are the same highest $M - m_i$ bids among $\mathbf{b}^{-i}$ as at the PNE. The selected losing bid of bidder n is not among them, so it remains losing. Therefore all m_i bids won by bidder i under $\mathbf{a}^i$ rank ahead of the selected bid.

This observation holds for every opponent. In the joint profile $(\mathbf{b}^n, \mathbf{a}^{-n})$, the opponents contribute $\sum_{i \neq n} m_i$ bids that rank ahead of the selected bid, and bidder n 's unchanged PNE-winning bids contribute another m_n . Hence

$$\sum_{i \neq n} m_i + m_n = \sum_i m_i = M, \quad (50)$$

so at least M bids rank ahead of every bid lowered in $\mathbf{b}_*^n$. Thus all of those bids are losing at the joint profile $(\mathbf{b}^n, \mathbf{a}^{-n})$. Lemma 2 can now be applied at this joint profile: replacing $\mathbf{b}^n$ by $\mathbf{b}_*^n$ lowers only bids that are losing there. Bidder n 's allocation is unchanged and the uniform price weakly decreases, so $u^n(\mathbf{b}_*^n, \mathbf{a}^{-n}) \geq u^n(\mathbf{b}^n, \mathbf{a}^{-n})$. Coupled with Definition 7, this means that $\mathbf{a}^{-n}$ is not a stay profile, a contradiction. This proves Equation (49).

Step 5: forgiving dynamics cannot converge to $\mathbf{b}$. Equation (49) is exactly the equilibrium condition of Theorem 4 with the substitute $\mathbf{b}_*^n$ from Equation (48). We conclude that forgiving GBA dynamics cannot converge to $\mathbf{b}$. $\square$

Making integer instances generic. Our experiments use integer values, which need not be generic: a bidder can be exactly indifferent between winning different numbers of units. Subtracting a tiny amount from every value fixes this. Fix $0 < \varepsilon < 1/M$ and set $\tilde{v}_m^n = v_m^n - \varepsilon$ for all n, m ; values stay positive and keep their order. Because bids are integers, the no-overbidding constraint $b_m^n < \tilde{v}_m^n$ allows exactly the same bids as $b_m^n < v_m^n$. So allocations and prices are unchanged at every bid profile; the only change is that each bidder's utility drops by ε per unit won.

The perturbed instance is generic. Fix a bidder and the other bidders' bids, and compare two of the bidder's bids that win $k \neq h$ units at uniform prices π and π' . Before the perturbation, the utility difference is an integer,

$$D = \left(\sum_{m=1}^k v_m^n - k\pi \right) - \left(\sum_{m=1}^h v_m^n - h\pi' \right) \in \mathbb{Z},$$

and afterwards it is $D - (k - h)\varepsilon$. Since $0 < |(k - h)\varepsilon| < 1$, this is not an integer, so it is not zero. The same calculation shows that the perturbation never reverses a strict preference; it only breaks exact ties, in favor of winning fewer units.

The perturbation also keeps non-spectating bidders non-spectating, because every bidder's highest feasible bid is unchanged. (Raising the opponents' bids can only cost a bidder units, so its highest bid guarantees it a unit exactly when it wins a unit against the opponents' highest bids.)

The perturbation barely changes the simulations. Fix the number of rounds T , the initial scores, and the Hedge parameters. Each payoff changes from $u_0^n(\mathbf{b})$ to

$$u_\varepsilon^n(\mathbf{b}) = u_0^n(\mathbf{b}) - \varepsilon x_n(\mathbf{b}),$$

where $x_n(\mathbf{b})$ is bidder n 's allocation. Each Hedge strategy is a continuous function of past expected payoffs, and these are continuous in ε and in earlier strategies. So, round by round, the play in each of the first T rounds converges to the unperturbed play as $\varepsilon \rightarrow 0$:

$$\max_{n, 1 \leq t \leq T} \|\mathbf{p}_{t,\varepsilon}^n - \mathbf{p}_{t,0}^n\|_1 \longrightarrow 0 \quad \text{as } \varepsilon \rightarrow 0.$$

At every bid profile, revenue is unchanged, while welfare and OPT both fall by exactly $M\varepsilon$. So every statistic we report (averages, extremes, and percentiles over our runs) converges to its unperturbed value as $\varepsilon \rightarrow 0$. In other words, a small enough ε gives a generic instance whose simulated outcomes match our reported integer-valued results to any desired accuracy.

E.4. FTL and PGD can remain at a positive-revenue UPA PNE

We now prove Proposition 7 in Section 6.3. In particular, we give a concrete positive-revenue UPA PNE and verify that weaker dynamics can remain there from the stated initializations. This also shows why the non-convergence result in Theorem 5 genuinely uses the small-gap exploration property.

PROPOSITION 7 (Weaker dynamics can remain at a positive-revenue UPA PNE). *There exists a UPA instance with a positive-revenue PNE at which MBA and no-regret learning dynamics can remain from suitable initializations.*

Proof. Consider a UPA with $N = M = 2$, where bidder 1 has priority in every tie. Let $\mathbf{v}^1 = \left(\frac{5}{2}, \frac{4}{3}\right)$, $\mathbf{v}^2 = \left(\frac{7}{3}, \frac{5}{4}\right)$, and consider the bid profile $\mathbf{b}^{1,*} = \mathbf{b}^{2,*} = (2, 1)$. The two bids equal to 2 win, so each bidder receives one unit. The highest rejected bid is 1, and hence the revenue is positive. This profile is a PNE. Against an opponent bidding $(2, 1)$, neither bidder can win both units because every feasible second bid is at most 1, whereas the opponent's first bid is 2. Any deviation that wins one unit pays the uniform price 1, while a deviation that wins no unit gives utility zero. Therefore bidder n can obtain utility at most $v_1^n - 1$, which is attained by $\mathbf{b}^{n,*}$. Thus $\mathbf{b}^*$ is a positive-revenue PNE. The instance also satisfies the auction-side conditions of Theorem 5 as each bidder owns one of the two highest feasible bids, and the values satisfy genericity.

FTL. For FTL, give each bidder a finite initialization bias U_0^n making $\mathbf{b}^{n,*}$ their unique cumulative-utility leader. Since $\mathbf{b}^{n,*}$ is a best response to $\mathbf{b}^{-n,*}$, its lead over every other bid never decreases. Hence FTL selects $\mathbf{b}^{n,*}$ at every round, and play remains at the PNE forever.

PGD. For PGD, initialize each bidder at $\mathbf{b}^{n,*}$. Since $\mathbf{b}^{n,*}$ is a best response to $\mathbf{b}^{-n,*}$, the update step given by Equations (11) and (14) leaves the agents' strategies—and thus their best responses—unchanged. By induction, PGD also remains at $\mathbf{b}^*$ forever. Thus MBA dynamics and no-regret dynamics can both remain at this positive-revenue PNE. $\square$