Auctions between Regret-Minimizing Agents

Yoav Kolumbus
The Hebrew University
Jerusalem, Israel
yoav.kolumbus@mail.huji.ac.il

Noam Nisan
The Hebrew University
Jerusalem, Israel
noam@cs.huji.ac.il

ABSTRACT

We analyze a scenario in which software agents implemented as regret-minimizing algorithms engage in a repeated auction on behalf of their users. We study first-price and second-price auctions, as well as their generalized versions (e.g., as those used for ad auctions). Using both theoretical analysis and simulations, we show that, surprisingly, in second-price auctions the players have incentives to misreport their true valuations to their own learning agents, while in first-price auctions it is a dominant strategy for all players to truthfully report their valuations to their agents.

CCS CONCEPTS
• Theory of computation → Algorithmic game theory and mechanism design: Convergence and learning in games.

KEYWORDS
Auctions, Regret Minimization, Repeated Games.

ACM Reference Format:
Yoav Kolumbus and Noam Nisan. 2022. Auctions between Regret-Minimizing Agents. In Proceedings of the ACM Web Conference 2022 (WWW ’22), April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3485447.3512055

1 INTRODUCTION

This paper deals with the following common type of scenario: several users engage in a repeated online auction, where each of them is assisted by a learning agent. A typical example is advertisers that compete for ad slots: typically, each of these advertisers enters his key parameters into a bidding-facing website, and then this website’s “agent” participates on the advertiser’s behalf in a sequence of auctions for ad slots. Often, the auction platform designer provides this agent as its advertiser-facing user interface. In cases where the platform’s agent does not optimize sufficiently well for the advertiser (but rather, say, for the auctioneer), one would expect some other company to provide a better (for the advertiser) agent. A typical learning algorithm of this type will have at its core some regret-minimization algorithm [11, 39], such as “multiplicative weights” (see [2] and references therein), “online gradient descent” [72], or some variant of fictitious play [14, 62], such as “follow the perturbed leader” [37, 45]. In particular, for ad auctions, there is empirical data showing that bids are largely consistent with no-regret learning [54, 60].

While one’s first intuition would be that an auction that is well designed for “regular” users would also work well for such agent-assisted users, it turns out that this is often not the case. In this paper we study how basic auction formats perform in terms of revenue and economic efficiency when software agents are doing the repeated bidding for the users. Specifically, we assume that each of the human bidders has a valuation into a regret-minimizing software agent, and then the agent engages in a long sequence of auctions, attempting to maximize utility for the human user. We present both simulation results and theoretical analysis for several types of auctions: first-price, second-price, and generalized first- and second-price auctions (for multiple ad slots). We study both classic regret-minimizing algorithms like multiplicative weights or follow the perturbed leader, as well as the consequences of the regret-minimization property itself.

Before presenting our results, let us say what exactly we are looking at when studying a long sequence of auctions played by software agents. While one might hope that there is some kind of convergence to fixed bids or distributions, it is well known that regret-minimization dynamics in many games need not converge [5, 6, 21, 40, 61]. For example, in our context, Figure 1a shows the dynamics of a simulation of a generalized first-price auction for two slots where the two bidders have the same value. (Throughout the dynamics, both agents bid nearly the same value and thus only a single bid-line is visible.) As one may clearly see, the dynamics do not converge. What we wish to study is the thing that is important to the auctioneer and to the bidders: the long-term distribution of the tuple of bids. For example, Figure 1b shows a density map of the empirical pair of bids in this simulation. The empirical distribution of the tuple of bids determines the average winning frequencies and prices, which in turn determine the users’ utilities and the auctioneer’s revenue over the whole sequence of auctions. While there may not always be a theoretical guarantee that this empirical distribution converges either, one may still theoretically analyze properties that it may have in various cases as well as study it through simulations.

It is well known that in any game and for any regret-minimizing dynamics the distribution of tuples of bids approaches a “coarse correlated equilibrium” [11, 37, 39, 71] (CCE), but as the space of CCEs is quite wide, this often does not suffice for an analysis of parameters of interest like the revenue or social welfare. For example, both for the first- and second-price auctions, there exist multiple (full information) CCEs, with different revenues and different social welfare [29]. One of the contributions of this work, presented in Section 4, is the identification of a refinement of the class of CCEs which we term co-undominated, to which alone a large class of regret-minimizing algorithms (“mean-based” ones, as defined in...
[13]) may converge in any game. In some cases this refinement suffices for an analysis, and for specific regret-minimization algorithms like multiplicative weights we may get an even finer analysis.

While there has been some work on the behavior of no-regret dynamics in auctions [7, 22, 23, 30], we do not know of previous analysis of the average utilities and revenue of standard no-regret dynamics in standard auction formats. The closest paper conceptually to ours is [30], which manages to analyze these for no-regret dynamics that are preceded by a long pure exploration phase by all agents. Here we provide theoretical analysis and simulations for standard regret-minimization algorithms (without a pre-exploration phase), sometimes getting very different results. In a companion paper [47] we discuss the convergence of regret-minimizing agents and analyze the incentives of the users of such agents in a general game-theoretic setting.

### 1.1 Second-Price Auctions

To demonstrate the difficulty, let us start by looking at the easiest-to-analyze auction: a second-price auction of a single item among two users. It is well known [28, 70] that such an auction is (dominant-strategy) incentive compatible, i.e., that no bidder ever has any incentive to misreport his value to the auctioneer. Such truthful bidding will result in an efficient allocation where the player with the higher value wins and pays a price equal to the value of the lower player.

So now let us look at our scenario, where the two (human) users each report their value to a software agent (one agent for each player), and then the two (software) agents repeatedly engage in auctions, each agent attempting to optimize for its own user according to the value reported to it. (We assume that each auction in the sequence is for the same type of item, and that the values of the players for this type of item are fixed and additive over the items won.) What do we expect to happen in this case? Specifically: Is it in the best interest of users to report their true values to their agents? Will the player with the higher value always win? Will he pay a price equal to the value of the lower player?

While positive answers to these three questions may seem to be trivially true, this turns out not to be the case, and only the second question gets a positive answer, while the answers to the first and third questions are negative.

**Theorem 1.** For two bidders with values \( v > w \) with multiplicative-weights agents that engage in a sequence of second-price auctions on their behalf with discrete bid levels, the distribution of bids of the high player converges to uniform in the range \((w, v]\), while the distribution of the bids of the low player converges to a distribution that has full support on \([0, w]\). In particular, in the limit, the high player always wins and pays an average price that is strictly less than the second price.

Figure 2a shows the dynamics of multiplicative-weights learning agents in a second-price auction, where the values of the two agents are \( v = 1 \) and \( w = 0.5 \), and Figure 2b shows the long-term distribution of bids of each of the two agents that indeed behave as the theorem states: the high-value agent always wins – except for a small number of auctions in the initial phase – and the average price he pays is about 0.27 (which is well below \( w \)). This is very much in contrast to the results of [30] that show that if we start with a sufficiently long pure exploration phase in which all agents bid uniformly, then dominant strategies are indeed reached. While the proved and observed lack of convergence to the dominant strategies may be surprising, the explanation is simple: while the dominant strategies do form a Nash equilibrium of the game, there are other (full-information) Nash equilibria of this game where the first player bids arbitrarily above the second price and the second player underbids arbitrarily.\(^1\) Regret dynamics do not necessarily choose the former and so we do get some distribution over the "low revenue" equilibria.

The fact that the average price paid is well below the second price has immediate implications for a human user that has a value that lies between the average price in these dynamics and the actual second price. Such a user will gain from misreporting (exaggerating) his value to his own agent. In this example, a player with value, say, \( v = 0.4 \) that bids truthfully against a second player with value \( w = 0.5 \) will always lose, whereas misreporting his value as \( v = 1 \) will lead to these simulated dynamics (as in Figure 2), making him always win and achieve a strictly positive utility on average. One can continue this analysis, viewing the reports of the users to their own agents as defining a "meta-game" between the users, and study the reached equilibrium. In a companion paper [47], we define and

---

\(^1\) These equilibria are distinct from the known "overbidding equilibria" where the low bidder bids above the high-bidders value and occur even if we completely disallow our learning agents to overbid as we do in our simulations.
whether the advertiser-optimizing software supplied by the search engines indeed optimizes for the advertiser (in the sense of regret minimization), and whether the search engines lose revenue by using such a generalized second-price auction.

1.2 First-Price Auctions

We now turn our attention to basic first-price auctions (for a single good). What do we expect to happen in a repeated first-price auction between two regret-minimizing algorithms with values \( v \geq w \)? Our basic intuition may be that they reach some kind of (full information) equilibrium of the first-price auction. It is well known that the only pure Nash equilibrium is where the low bidder bids his value (up to an \( \epsilon \)) and the high bidder bids \( \epsilon \) more than that. This means that we would expect the high player to win (almost always) and pay (close to) the second price. Mixed Nash and correlated equilibria of the first-price auction are also known to yield the same outcome (namely, the high player always wins and pays the second price) [29]. Figure 3 strengthens this intuition by showing the dynamics of first-price auctions with values \( v = 1 \) and \( w = 0.5 \) (Figure 3a), and with symmetric values \( v = w = 1 \) (Figure 3b), between two multiplicative-weights algorithms that indeed converge to this outcome.

The missing element in this intuition is that regret-minimization dynamics need not converge to a Nash equilibrium (or even a correlated equilibrium) but may reach an arbitrary coarse correlated equilibrium, and it turns out that the first-price auction also has other coarse correlated equilibria [29], and in many of these coarse correlated equilibria the revenue is well below the second price, and sometimes even the low player wins. While in [29] these coarse correlated equilibria were analyzed in a continuous setting, they also exist with simple discrete bidding levels. For example, consider an auction with \( v = w = 1 \); the reader may verify that the distribution that gives equal weights to the following four pairs of bids is a coarse correlated equilibrium: \((0,0),(0.46,0.46),(0.64,0.64),(0.73,0.73)\).

It is true in general that for every CCE there exist some regret-minimizing algorithms that converge to it ([53], see also [47]). Specifically, in our setting, consider agents that start by playing a sequence of bids that converge to the desired CCE (according to some fixed-in-advance schedule); each software agent proceeds this way as long as the other agent does, and in the case of any deviation reverts to a standard regret-minimizing algorithm. Thus, regret-minimizing agents need not converge to the "second-price outcome" in general. We are able to prove, however, that a large class of regret-minimizing algorithms, defined in [13] and called "mean-based" algorithms, may converge only to this second-price outcome. Informally, algorithms from this class cannot put too much probability on actions that previously performed poorly.

**Theorem 2.** In a repeated first-price auction between two mean-based regret-minimizing agents with possible bid levels that are a discrete \( \epsilon \)-grid, if the dynamics converge to any single distribution, then the high player wins with probability approaching 1 and pays an average price that converges to the second price (up to \( O(\epsilon) \)).

Following our simulations, we conjecture that the assumption that the dynamics converge (a notion defined in Section 3) is not really required, at least for the multiplicative-weights algorithm.

As opposed to the case of the second-price auction where our results were in contrast to those with an additional pre-exploration phase [30], here we get the same type of result but for natural algorithms, without requiring this additional exploration phase.

A rather direct corollary of the fact that we get the second-price outcome is that from the point of view of the human users that report their values to their regret-minimizing agents, they observe a second-price auction. Since this auction is incentive compatible, we get that there is never any incentive for a user to misreport his value to his own agent (up to a vanishing loss in the long term). Recall that this is in contrast to the case of the second-price auction described above.

1.3 Generalized First-Price Auctions

We now proceed to the more complex generalized first-price auction. Specifically, we focus our attention on the following first price two-slot auction: two “ad slots” are being sold in an auction to two bidders. The "top" ad slot has a "click-through rate" of 1, while the "bottom" ad slot has a click-through rate of 0.5. Each of the bidders has a value "per click" and so his value for the bottom slot is exactly half of his value for the top slot.

Our first result is the analysis of the (mixed) Nash equilibrium of this game. While there has been much work on equilibria in related auction settings (e.g., [19, 27, 29, 42]), to the best of our knowledge, this paper is the first to successfully analyze the Nash equilibrium of a generalized first-price auction.

**Theorem 3.** Assume w.l.o.g. that \( v \geq w \). In the unique Nash equilibrium of the two-slot auction described above, the players mix...
The expected payoffs are: $u_1 = v/2 + (v-w)(1 - \ln 2)$, and $u_2 = w/2$.

Running an actual simulation of multiplicative-weights agents participating in such an auction with $v = w = 1$ shows that the agents do not get even close to any uncoupled mixing profile as this Nash equilibrium. Figure 1a shows the dynamics of a (typical) simulation. Throughout the dynamics, both agents bid nearly the same and thus only a single bid-line is visible. Figure 1b shows the density of the empirical pair of bids. The lack of convergence to a Nash equilibrium is not surprising since we only expect to reach a CCE of the game, not necessarily the Nash equilibrium.

Most of Section 7 is devoted to studying the equilibrium that is actually reached by multiplicative-weights algorithms. We observe via simulations that the bid distribution of the high bidder agrees with the computed distribution of the Nash equilibrium $F(x) = x / w - x$, but that the other player’s bid is strongly correlated with the high player’s bid (rather than independent of it as in the Nash equilibrium). In fact, we see that in the achieved equilibrium the low player either bids (almost) identically to the first player or bids close to 0, where the probability of the latter is proportional to the ratio between the high and low bids. As we show in Section 7, our closed-form formulas agree well with the simulation results, but we do not have a complete theoretical analysis of this and such an analysis seems to be difficult.

The closed-form formulas do allow us, however, to analytically compute the achieved winning probabilities and prices in this auction and thus also the users’ payoffs from different reports in the “meta-game” where the users report their values to their respective agents. In particular, we see that even with true reports, the users’ average utilities are (slightly) higher than in the Nash equilibrium, implying that their agents achieved some sort of implicit collusion. Demonstrating an analysis of strategic reports of the users to their agents, Figures 4a and 4b show the utilities that users with a true value of 1 would get from strategically reporting different values to their own agent as a reply to two fixed reports ($0.54$ and 2)$^5$ of the other user. We first see the very close agreement of the closed-form formulas with the actual simulation results. Second, these results demonstrate the equilibrium in the meta-game where one bidder exaggerates his bid as much as possible, while the other one underbids somewhat. Calculating the utilities of the players in this meta-game equilibrium reveals that severe implicit collusion occurs, where the users capture 89% of the total welfare, leaving only 11% to the auctioneer (in contrast to 33% in the Nash equilibrium and 31% in the equilibrium reached for truthful reports).

2 FURTHER RELATED WORK

Regret minimization in repeated strategic interactions and in online decision problems has been extensively studied in the literatures of game theory, machine learning, and optimization. Early regret-minimizing algorithms were related to the notion of fictitious play [14, 62] (a.k.a. “follow the leader”), which in its basic form does not guarantee low regret, but whose smoothed variants, such as “follow the perturbed leader” (FTPL) [35, 37, 44, 45] or “follow the regularized leader” (FTLR) [66], are known to guarantee adversarial regret of $O(\sqrt{T})$ in $T$ decisions. Another common approach to regret minimization is the “multiplicative-weights” algorithm which has been developed and studied in many variants (see [2] and references therein). Similar ideas of regret minimization apply both in settings in which players observe full feedback on their past actions and forgone actions (the experts-advice setting) and in the multi-armed bandit setting, where feedback is available at every step only for the action that was selected at that step, and in which variants of multiplicative weights (such as EXP3 [3, 49]) and other algorithms have been developed.

Regret minimization has also been widely studied in auctions, with two prominent lines of work focusing on the econometric problem of inferring structural parameters of the utilities of the bidders [33, 54, 58, 59], and on the price of anarchy in classic auction formats and in combinatorial auctions when the bidders are no-regret learners [10, 16, 22, 41, 63, 64, 67]. Other works deal with the design of automated bidding algorithms for auctions with more complex factors, such as budget constraints [7, 24], or the utilization of regret minimization ideas for optimizing reserve prices [17, 52, 65], and in [1, 60] regret-minimization algorithms were used to predict bidder behavior. Importantly, [54, 60] provide concrete empirical evidence from large ad auction datasets showing that actual bidder behavior is consistent with no-regret learning.

The standard first-price and second-price auctions have been widely studied in auction theory [48, 51]. Their generalized versions for multiple items have been studied and applied in ad auctions (also called “position auctions”) [12, 28, 43, 69]. Most prior works mainly discussed the Nash equilibria of the auctions or their best-response dynamics, and did not study the dynamics when players use regret-minimization algorithms. One notable exception is [30]; however, as discussed in the introduction, they analyze a different setting in which the actual play of the agents is preceded by a pure exploration phase in which all bidders bid uniformly, which leads to different convergence results than those we show for standard interaction between the algorithms.

Finally, there is a growing body of work studying various aspects of interactions between strategic players and learning algorithms, including learning from strategic data [15, 20, 26, 34, 36, 38, 50], Stackelberg games [9, 32], security games [31, 55, 56], recommendation systems [8, 68], and optimization against regret-minimizing
opponents [13, 25]. Notably, we use the formalism of mean-based learning algorithms defined in [13], who also study an auction setting. Their setting, however, is basically different from ours. They consider a game in which a strategic seller seeks revenue-maximizing auction strategies against a regret-minimizing buyer, while we consider fixed auction rules and study the interactions between regret-minimizing buyers and the strategic considerations of their users. Lastly, in a companion paper [47], we formalize a “meta-game” model for general strategic settings in which human users select parameters to input into their learning agents that play a repeated game on their behalf, and we analyze this model for simple games with general regret-minimizing agents. Our discussion in the present paper on the incentives of players to misreport their valuation to their own bidding agents is directly related to this meta-game model.

3 PRELIMINARIES

We consider repeated auction settings in which the same group of bidding agents repeatedly play an auction game with a fixed auction rule and the same set of items are being sold at each repetition of the game. The valuation of every agent, assigned to it by its user, is constant throughout the dynamics, and the utilities are additive. i.e., a bidder with value \( v \) for an item who wins \( k \) such items has total value \( kv \). Player values and bids are assumed to be on an \( \epsilon \)-grid, i.e., to be multiples of a minimum value \( \epsilon > 0 \) (e.g., 1 cent). We assume that agents cannot overbid, and in all auction formats if agents place equal bids, ties are broken uniformly at random.

First-price and second-price auctions are defined in the standard way: the auction sells a single item, and \( b \) 1 wins the auction, and pays a price \( p \) to the auctioneer. In the first-price auction the winner pays his bid: \( p = b_1 \), and in the second-price auction the winner pays the amount of the next highest bid: \( p = b_2 \). The utility for player \( s \) with value \( v_s \) is then \( v_s - p \) when winning the item, and zero otherwise. For further details on these auction formats see, e.g., [57]. Generalized auctions for multiple items are defined in the relevant sections.

We study agents that use regret-minimization algorithms. The regret of player \( i \) at time \( T \), given a history of bids \( (b_1, \ldots, b_T) \), is defined as the difference between the optimal utility from using a fixed bid in hindsight and the actual utility: \( R_i^T = \max_b \sum_{t=1}^T u_i(b, b_t) - u_i(b_i^T, b_t^*) \), where \( b_i^T \) is the bid of player \( i \) at time \( t \) and \( b_t^* \) denotes the bids of the other players at time \( t \). As usual, regret-minimization algorithms are stochastic, and whenever we talk about the limit behavior, we consider a sequence of algorithms with \( T \to \infty \) and with probability approaching 1. An agent \( i \) is said to be “regret-minimizing” if \( R_i^T / T \to 0 \) almost surely as \( T \to \infty \). In our simulations we mainly present the multiplicative-weights algorithm, as defined in [2]. The results apply equally both to the Hedge algorithm (often also referred to as multiplicative weights in the literature) and to the linear version of the algorithm. Appendix A presents simulations with additional algorithms.

We use the following notations to describe the empirical dynamics of the agents’ bids. Denote by \( \Delta \) the space of probability distributions over bid tuples in an auction with discrete bid levels. We use \( p_t^j \in \Delta \) to denote the empirical distribution of bids after \( t \) rounds in a sequence of \( T \geq t \) auctions, and denote by \( p_t^j(b) \) the empirical frequency of a bid tuple \( b \) after round \( t \). When we discuss convergence of the dynamics we consider the following definition:

**Definition 1.** The dynamics “converge to a distribution” \( p \in \Delta \) if for every \( \epsilon > 0 \) there exists \( T_0(\epsilon) \) s.t. for every \( T > T_0 \) w.p. at least \( 1 - \epsilon \) it holds that for every \( \epsilon T < t \leq T \), \( |p_t^j - p| < \epsilon \).

Intuitively, this means that after a sufficiently long time, the time-average distribution stabilizes and remains close to a stationary distribution. Notice that the convergence of the cumulative empirical distribution to a joint distribution does not imply the convergence of the probability of play of any single agent to a distribution on its own actions (a stronger notion of convergence studied, e.g., in \( [6, 23, 30] \)). See also [47] for a discussion of yet weaker notions of convergence.

4 CO-UNDOMINATED COARSE EQUILIBRIA

In this section we define a refinement of the concept of coarse correlated equilibria, which we call “co-undominated CCE.” For simplicity, the definitions are given for two-player games; their generalization is straightforward. Intuitively, those are CCEs in which players do not use strategies that are weakly dominated relative to the distributions of the opponents. Formally:

**Definition 2.** Let \( A, B \) denote subsets of the action spaces of the two players in a two-player game. Action \( i \in A \) of player 1 is called weakly-dominated-in-B by action \( i' \) of player 1 if \( \forall j \in B : u_i(i, j) \leq u_i(i', j) \), and \( \exists j \in B : u_i(i, j) < u_i(i', j) \).

**Definition 3.** Let \( p_{i,j} \) be a CCE of a (finite) two-player game with action spaces \( I, J \). Denote its support (\( A, B \)) by \( A = \{ i \in I \mid \exists j \in J \text{ with } p_{i,j} > 0 \} \) and \( B = \{ j \in J \mid \exists i \in I \text{ with } p_{i,j} > 0 \} \) (\( A \) is a subset of \( I \) and \( B \) is a subset of \( J \)). A CCE is called co-undominated if no strategy in its support is weakly dominated relative to the support of the other player, i.e., if for every \( i \in A \) and every \( i' \in I \), action \( i \) is not weakly-dominated-in-B by \( i' \), and similarly for \( B \).

We show that these types of equilibria are closely connected to a family of learning algorithms called “mean-based.” For completeness, we cite their definition (\[13\], Definition 2.1):

**Definition 4.** (Mean-based learning algorithm, \[13\].) Let \( \sigma_{i,t} = \sum_{i'=1}^{I} u_{i,i'} \), where \( u_{i,t} \) is the utility of action \( i \) at time \( t \). An algorithm for the experts problem or the multi-armed bandits problem is \( \gamma(T) \)-mean-based if it is the case that whenever \( \sigma_{i,t} < \sigma_{i',t} - \gamma(T)T \), then the probability that the algorithm pulls arm \( i \) in round \( t \) is at most \( \gamma(T) \). We say an algorithm is mean-based if it is \( \gamma(T) \)-mean-based for some \( \gamma(T) = o(1) \).
The above definition from [13] considers a fixed value $T$. To clarify the definition further when considering a series of $T$-values and the limit $T \to \infty$, we say that an algorithm family $A^T$, parametrized by $T$, is mean-based if there exists a series $\gamma(T)_{T=1}^{\infty}$ such that $\gamma(T) \to 0$ as $T \to \infty$, and each algorithm in the series $(A^T)_{T=1}^{\infty}$ is $\gamma(T)$-mean-based.

The following lemma connects between mean-based regret-minimization algorithms and co-undominated CCEs. This result is general for any game (not only auctions), and will be useful for our analysis in the following sections. The proof is in Appendix B.

**Lemma 1.** Consider the dynamics of any mean-based regret-minimizing agents playing a repeated finite two-player game. If the dynamics converge to a distribution $p$, then $p$ is a co-undominated coarse correlated equilibrium.

### 5 SECOND-PRICE AND GSP AUCTIONS

We consider agents that play repeated second-price auctions (for a single item) and GSP auctions (for multiple items; see, e.g., [28]). Our analysis starts with the simple second-price auction, which is known to be incentive compatible; i.e., it is a dominant strategy for all players to bid their true values. We show that in the setting where regret-minimizing agents play the auction on behalf of their users, the results are different from this truthful equilibrium.

Specifically, Theorem 1 shows that when the auction is played between two multiplicative-weights agents, in the limit, although the high player always wins the auction (as in the truthful equilibrium), he pays a price that is strictly less than the second price. Before proceeding to the proof, we consider the following technical restatement of the theorem.

**Theorem.** *(Restatement of Theorem 1):* In the repeated second-price auction with discrete bid levels that are multiples of $\epsilon$ and players using multiplicative-weights algorithms with values $v > w$, in the limit the dynamics converge to a joint distribution in which for the "high player" with value $v$, $Pr[0] = Pr[\epsilon] = \ldots = Pr[w] = 0$, and $Pr[w + \epsilon] = Pr[w + 2\epsilon] = \ldots = Pr[v] = \frac{v-w}{v-w}$, and for the "low player" with value $w$, $0 < Pr[0] \leq Pr[\epsilon] \leq \ldots \leq Pr[w]$.

**Proof.** First, observe that for the low player, bidding $w$ weakly dominates all other bids, and so the probability assigned to this bid does not decrease in any update of the weights. Thus, the probability that the low player bids exactly $w$ is at least\(^8\) $1/(w/\epsilon + 1)$.

Since the low player bids $w$ at least $1/(w/\epsilon + 1) \times (w/\epsilon)$ of the time (with high probability), every bid that is less than $w$ of player 1 loses the auction at this frequency, while an alternative bid of $w + \epsilon$ always wins, yielding positive utility. Thus, every $O(w/\epsilon)$ steps the probability that the high player bids $w$ or less decreases at least by a constant factor.

Hence, we can deduce that the expected number of times that the high player bids below $w$ is a constant, since it is bounded by the sum of a geometric series. In light of this observation, we can look at the bid distribution of the low player. Whenever the high player bids above $w$, the weights of all bids do not change (since all bids would yield zero utility). The weights are updated only in those cases where the high player bids below $w$. Since this happens only a finite number of times, the ratio of probabilities between any two bids of the low player remains bounded by a constant. Thus, every bid of the low player remains with positive probability in the limit (and at any time). Since whenever a bid $j$ of the low player would win the auction, every bid $k > j$ would also win the auction, it follows that $Pr[0] \leq Pr[\epsilon] \leq \ldots \leq Pr[w]$.

Finally, since all bids above $w$ have the same utility at every step (as the payment is the second price), all weight updates of every bid above $w$ are exactly equal, and so the distribution of bids of the high player in the limit is uniform between $w + \epsilon$ and $v$.

In addition to the above result that explicitly shows that dynamics of multiplicative-weights agents converge to these forms of distributions, our analysis from the previous section allows us to show that for two agents implemented as any mean-based regret minimization algorithms, whenever the dynamics of the agents converge to a coarse correlated equilibrium, it must be an equilibrium in which the agent with the higher valuation always wins. The proof of the following lemma is in Appendix C.

**Lemma 2.** In the repeated single-item second-price auction with discrete bid levels that are multiples of $\epsilon$ and two players with values $v > w$, in any co-undominated CCE the player with the higher valuation $v$ always wins the auction.

These theoretical results are consistent with simulation results, as can be seen in Figure 2 that shows the bid dynamics of two multiplicative-weights agents playing a repeated second-price auction for a single item. The first player has value 1 for the item and the second player has value 0.5. In the very beginning, bids are close to uniform and there are some auctions in which player 2 bids higher than player 1 and wins the item with positive utility. However, after a short learning phase player 1 learns to bid higher than the value 0.5 of player 2 and to always win the auction.

Figure 2b shows the empirical bid densities of the two players. It can be seen that the high player bids uniformly between his value and the value of the low player. The basic reason for this is that under the second-price payment rule, all these bids are equivalent. The bids of player 2 are not uniform, but have more weight on higher bids. This is despite the fact that in the long-term dynamics player 2 always loses, getting zero utility, and so all his bids should also be equivalent. The reason for this form of the distribution is that the multiplicative-weights algorithm (and other algorithms like FTPL) keep a "memory" of the history of play: in the initial

---

\(^8\)For simplicity, we use the standard definition of the algorithm where the weights are initialized uniformly. The result that the high player always wins and pays strictly less than the second price extends to any other initialization that has full support.
phase there was a finite number of auctions in which player 2 got positive utility. These winnings were associated with higher bids. Thus, in the initial phase higher bids get higher weight, and after this phase all bids yield exactly the same feedback (of zero utility), and therefore the distribution remains skewed toward the higher bids. This intuition is formalized in the proof of Theorem 1 above.

Figure 5a shows a very similar picture also in the generalized second-price auction with multiple items. Interestingly, there is a narrow bid range (centered at the bid 0.4 in Figure 5b) in which the bid supports of both bidders overlap, but, as can be seen by the distribution of bid differences at all times (in green), the bids are correlated such that despite the overlapping bid ranges, the player with the higher valuation still almost always wins the auction.

Now let us consider the point of view of the human users that report their values to their regret-minimizing agents playing a repeated second-price auction on their behalf. Although this auction is known to be incentive compatible, in light of the results shown above on the dynamics of the agents, from the point of view of the human users the second-price outcomes are not obtained, and thus these users do not observe a second-price auction. Hence, in the repeated second-price auction that is implemented with regret-minimizing agents it can be beneficial for players to misreport their true valuations.

To demonstrate this, let us return to our example from the introduction in which two players compete in a repeated second-price auction using multiplicative-weights agents. The first player has value $0.4$ and the second player has value $0.5$. Figure 2c shows the expected utility for player 1 over a period of 50,000 auctions from inputting different values $v$ into his agent. It can be seen that when $v$ is sufficiently high above the value $0.5$ of player 2, the utility of player 1 is positive and increases for yet higher values; i.e., for a fixed value of the low player, the high player pays a lower price as he declares a higher value. For very high values of the high player, the high player’s agent almost never bids below the low value, and so the distribution of bids of the low player remains very close to uniform. As the high declaration decreases, the probability that the high player’s agent bids below the low value increases, and this is what drives the increase in the probability of the low player’s agent bidding higher values.

It follows that the best reply of the high player in the meta-game is always as high a declaration as possible, as long as the low player’s declaration is below a maximum value that is less than twice the true value of the high player, and under which the dynamics lead to a positive average utility for the high player. The low player (almost) always loses and so all declarations that cause him to lose are equivalent. These can lie anywhere between 0 and the declaration of the high player. The truthful declaration is certainly one possible best reply in the full information game, but notice that there is little to suggest selecting it specifically since it is a best reply only if it leads to a loss, and is in no way a dominant strategy if the value of the other player is not known. We thus can identify a whole set of possible equilibria of the meta-game, but do not see a natural way to select one of these as the outcome of user behavior in the meta-game. Notice that the revenue is close to half of the low declared value (when the high declaration is very high) and thus we cannot give a reasonable prediction of the revenue in equilibrium.

6 \textbf{FIRST-PRICE AUCTIONS}

In this section we present our analysis of first-price auctions leading to Theorem 2. We begin by showing that in the first-price auction in any co-undominated CCE, all bids of the high-value player are close to the second price. This is formalized in the following lemma. The proof of the lemma is deferred to Appendix D.

\textbf{Lemma 3.} In a first-price auction with discrete bid levels that are multiples of $\epsilon$ with player values $v \geq w > 2\epsilon$, in any co-undominated CCE, the support of the high player with value $v$ is in $[w-2\epsilon, \ldots, w+\epsilon]$. Additionally, when $\epsilon$ is small compared to the value difference $v-w$, the player with the high value $v$ almost always wins the auction.

Using this result together with Lemma 1 that connects between mean-based learning agents and co-undominated CCEs, we can readily prove Theorem 2:

\textbf{Proof. (Theorem 2):} Consider a first-price auction with discrete bid levels that are multiples of $\epsilon$ with players with values $v \geq w > 2\epsilon$ that repeatedly play the auction using mean-based regret minimization algorithms. By Lemma 1, if the dynamics converge to a CCE, then it must be a co-undominated CCE. By Lemma 3, in the first-price auction in such an equilibrium both players bid close to the low value $w$ (up to $2\epsilon$), and the player with the higher value $v$ wins with probability approaching $1$ as $\epsilon/(v-w) \to 0$. \hfill $\Box$

7 \textbf{GENERALIZED FIRST-PRICE AUCTIONS}

We analyze GFP auctions where two “ad slots” are being sold to two bidders. The “top” ad slot has a “click-through rate” of $1$, while the “bottom” ad slot has a click-through rate of $0.5$. That is, a bidder with value $v$ that wins the top slot and pays $p$ (per click) gets a utility of $v-p$, and the bidder that gets the bottom slot and pays $p$ (per click) gets a utility of $v/2-p/2$. The auction uses the first-price rule: the bidder that bids highest gets the top slot and the low bidder gets the bottom slot, and each pays their bid per click. That is, if the first player has value $v$ and bids $x$ while the second player has value $w$ and bids $y$, then the utility of the first player is $v-x$ if $x>y$ and is $(v-x)/2$ if $x<y$, while the utility of the second player is $w-y$ if $y>x$ and is $(w-y)/2$ if $y<x$. (In the case of a tie, $x=y$, let us assume that we toss a coin for the top slot, in which case the utility of the first bidder is $3 \cdot (v-x)/4$, and the utility of the second bidder is $3 \cdot (w-y)/4$.

Our first result on this auction is the characterization of its unique Nash equilibrium, which is specified in Theorem 3. The proof of the theorem is deferred to Appendix E. Next, we proceed to analyze the empirically observed distributions obtained by the agents that repeatedly play this auction.

7.1 \textbf{Coarse Correlated Equilibria}

As discussed in the introduction, the dynamics of multiplicative-weights agents with equal values $v = w = 1$ yield, on average, the Nash equilibrium distribution as stated in Theorem 3; however, unlike in the mixed equilibrium, here bids are highly correlated. Throughout the dynamics, as shown in Figure 1a, both agents bid nearly the same, and the bids do not converge. The lack of convergence to a Nash equilibrium is not surprising since we expect to reach only a CCE of the game, and not necessarily the Nash equilibrium. So let us analyze the empirically selected CCE.
utilizes of the players to those implied by the Nash equilibrium and to those implied by our prediction. As we see, while the observed empirical utilities are very far from those in the Nash equilibrium, they agree with our prediction quite well.

7.2 Equilibria of the Meta-game

We now proceed to study the "meta-game" [47] faced by the users, each of whom needs to report his value to his own multiplicative-weights regret-minimization algorithm. Our analysis concerns the case where the true valuations of the players are 1 and 1 and their declarations to their agents are $v$ and $w$. Assuming that the agents reach the "equilibrium" predicted by our nearly diagonal distribution model, we can calculate the players’ utilities. Our analysis allows us then to identify the ($\epsilon$-)Nash equilibria of the meta-game. The formal details of this analysis can be found in the full version of the paper [46]. In the equilibrium of the meta-game, the high player declares an arbitrarily high value $v$ and the low player declares a fixed value $w = 1/(6 \cdot (1 - \ln(2))) \approx 0.54$, and the utilities that they get, respectively, in this equilibrium are $5/6 - O(1/v)$ and $1/2 + O(1/v)$. (If there is a maximum allowed value for $v$ this is exactly an equilibrium, and otherwise it is an $\epsilon$-equilibrium, where $\epsilon = O(1/v)$.) The interesting point of this analysis is that the players are locked in a type of hawk-dove game: there are two possible equilibrium points depending on which player is the high bidder and which is the low one. Each player prefers to be the high player and to let the other player be the low player. However, if they both declare high values, then they both suffer and get negative utility.

This theoretically predicted behavior is indeed borne out in simulations. Figure 4a shows the average payoffs of the two players when player 1 uses a fixed "high" declaration $v = 2$ as a function of the declaration $w$ of the other player. Figure 4b shows the payoffs when player 2 uses the equilibrium declaration $w = 1/(6 \cdot (1 - \ln(2)))$ as a function of the declaration $v$ of player 1. It can be seen that the nearly diagonal distribution model is consistent with the empirical payoffs, and that the predicted equilibrium of the meta-game is indeed empirically an equilibrium as the best replies of the players behave as predicted: the best-reply of the "low player" is obtained at $w \approx 0.54$ (marked in red in Figure 4a), and the payoff of the "high player" is increasing in the declaration $v$, approaching $5/6 \approx 0.83$ for high declaration values (Figure 4b). These results are robust for different player parameters and several other learning algorithms such as linear or exponential multiplicative weights, FTPL, and FTRL with a recency bias.

To summarize, we succeeded to identify (and empirically validate) the outcome of dynamics of multiplicative-weights agents in this auction with any two values. This allowed us to analyze the equilibrium in the meta-game induced on users that have identical true values. We found two such equilibria where, in each of them, one player takes the role of the high player, declaring the maximum possible value to his agent, and the other player declares a value of 0.54 fraction of the true value. Significantly, in each of these equilibrium points the users capture $8/9 \approx 89\%$ of the welfare, leaving only 11% to the auctioneer. This outcome exhibits significant implied collusion between the users and, in fact, suggests that such an auction platform encourages collusion, probably against the interests of the platform designer.
ACKNOWLEDGMENTS
This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (grant agreement no. 740282).

REFERENCES
[1] Saeed Alaee, Adhwinikumar Badanidiyuru, Mohammad Mahdian, and Sadra Yazdanbod. 2019. Response prediction for low-regret agents. In International Conference on Web and Internet Economics. Springer, 31–44.
[2] Sanjeev Arora, Elad Hazan, and Satyen Kale. 2012. The multiplicative weights update method: a meta-algorithm and applications. Theory of Computing 8, 1 (2012), 121–164.
[3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-time analysis of the multiarmed bandit problem. Machine learning 47, 2 (2002), 235–256.
[4] Peter Auer, Nicolò Cesa-Bianchi, and Claudio Gentile. 2002. Adaptive and self-confident on-line learning algorithms. J. Comput. System Sci. 64, 1 (2002), 48–75.
[5] James P. Bailey, Sai Ganesh Nagarajan, and Georgios Piliouras. 2021. Stochastic Multiplicative Weights Update in Zero-Sum Games. arXiv preprint arXiv:2110.02134 (2021).
[6] James P. Bailey and Georgios Piliouras. 2018. Multiplicative weights update in zero-sum games. In Proceedings of the 2018 ACM Conference on Economics and Computation. 321–338.
[7] Santiago Ballester and Yanontan Gür. 2019. Learning in repeated auctions with bandits: Regret minimization and equilibrium. Management Science 65, 9 (2019).
[8] Omer Ben-Porat and Moshe Tennenholtz. 2018. A game-theoretic approach to recommendation systems with strategic content providers. In Proceedings of the 32nd International Conference on Neural Information Processing Systems. 1118–1128.
[9] Georgios Birmpas, Jiariu Gan, Alexandros Hollender, Francisco J Marmolejo-Cossio, Ninad Rajgopal, and Alexandros A Voudouris. 2020. Optimaly Deceiving a Learning Leader in Stackelberg Games. arXiv:2006.05666 (2020).
[10] Avrim Blum, MohammadTaghi Hajiaghayi, Katrina Ligett, and Aaron Roth. 2008. Regret minimization and the price of anarchy. In Proceedings of the forty-ninth annual ACM symposium on Theory of computing. 373–382.
[11] Avrim Blum and Tishyaa Monsour. 2007. Learning, regret minimization, and equilibria. (2007).
[12] Christian Borgs, Jennifer Chayes, Nicole Immorlica, Kamal Jain, Omid Etesami, and Mohammad Mahdian. 2007. Dynamics of bid optimization in online advertisement auctions. In Proceedings of the 16th international conference on World Wide Web. 531–540.
[13] Mark Braverman, Jierning Mao, Jon Schneider, and Matt Weinberg. 2018. Selling to a no-regret buyer. In Proceedings of the 2018 ACM Conference on Economics and Computation. 523–538.
[14] George W. Brown. 1951. Iterative solutions of games by fictitious play. Activity analysis of production and allocation 13, 1 (1951), 374–376.
[15] Yang Cai, Constantinos Daskalakis, and Christos Papadimitriou. 2015. Optimum statistical estimation with strategic data sources. In Conference on Learning Theory. PMLR, 280–296.
[16] Ioannis Caragiannis, Xiannan Ghanne, Vinfu Ndam Biay, Inbal Talgam-Cohen, and Nir Rosenfeld. 2021. Strategic Classification in the Dark. J. Comput. 139 (2021), 3672–3681. https://proceedings.mlr.press/v139/ghanne21a.html.
[17] Michael D Grgoriadis and Leonid G Khachian. 1995. A sublinear-time randomized approximation algorithm for matrix games. Operations Research Letters 18, 2 (1995), 53–58.
[18] Nika Haghtalab, Nicole Immorlica, Brendon Lucier, and Jack Z. Wang. 2021. Maximizing Welfare with Incentive-Aware Evaluation Mechanisms. , Article 23 (2021), 7 pages.
[19] James Hannah. 1957. Approximation to Bayes Risk in Repeated Play. In Contributions to the Theory of Games (AM-39), Volume III. Princeton University Press, 97–139.
[20] Mostra Hardt, Nimrod Megiddo, Christos Papadimitriou, and Mary Wootters. 2016. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science. 111–122.
[21] Sergiu Hart and Andreu Mas-Colell. 2009. A simple adaptive procedure leading to correlated equilibrium. Econometrica 68, 5 (2000), 1127–1150.
[22] Sergiu Hart and Andraus Mas-Colell. 2013. Simple adaptive strategies: from regret-matching to uncoupled dynamics. Vol. 4. World Scientific.
[23] Jason Hartline, Vasileios Syrgkanis, and Eva Tardos. 2015. No-Regret Learning in Bayesian Games. Advances in Neural Information Processing Systems 28 (2015), 3061–3069.
[24] Yoav Kolumbus and Noam Nisan. 2021. Auctions Between Regret-Minimizing Agents. arXiv preprint arXiv:2110.11855 (2021). https://arxiv.org/pdf/2110.11855.pdf.
[25] Yoav Kolumbus and Noam Nisan. 2021. How and Why to Manipulate Your Own Agent. arXiv preprint arXiv:2112.07640 (2021). https://arxiv.org/pdf/2112.07640.pdf.
[26] Vijay Krishna. 2009. Auction theory. Academic press.
[27] John Langford and Tong Zhang. 2007. Epoch-Greedy algorithm for multi-armed bandits with side information. Advances in Neural Information Processing Systems (NIPS 2007) 20 (2007), 1.
[28] Paul Milgram et al. 1987. Auction theory. In Advances in economic theory: Fifth world congress, Vol. 1. Cambridge University Press Cambridge, 32.
[29] Dambule Monot and Georgios Piliouras. 2017. Limits and limitations of no-regret learning in games. The Knowledge Engineering Review 32 (2017).
[54] Denis Nekipelov, Vasilis Syrgkanis, and Eva Tardos. 2015. Econometrics for Learning Agents. In Proceedings of the Sixteenth ACM Conference on Economics and Computation (Portland, Oregon, USA) (EC ’15). Association for Computing Machinery, New York, NY, USA, 1–18. https://doi.org/10.1145/2764468.2764522

[55] Thanh Nguyen and Haifeng Xu. 2019. Imitative Attacker Deception in Stackelberg Security Games. In IJCAI 528–534.

[56] Thanh H Nguyen, Yongzhao Wang, Arunesh Sinha, and Michael P Wellman. 2019. Deception in finitely repeated security games. In Proceedings of the AAAI Conference on Artificial Intelligence.

[57] Noam Nisan et al. 2007. Introduction to mechanism design (for computer scientists). Algorithmic game theory 9 (2007), 209–242.

[58] Noam Nisan and Gali Noti. 2017. An Experimental Evaluation of Regret-Based Econometrics. In Proceedings of the 26th International Conference on World Wide Web (Perth, Australia) (WWW ’17). International World Wide Web Conferences Steering Committee. Republic and Canton of Geneva, CHE, 73–81. https://doi.org/10.1145/3038912.3052621

[59] Noam Nisan and Gali Noti. 2017. A "Quantal Regret" Method for Structural Econometrics in Repeated Games. In Proceedings of the 2017 ACM Conference on Economics and Computation (Cambridge, Massachusetts, USA) (EC ’17). Association for Computing Machinery, New York, NY, USA, 123. https://doi.org/10.1145/3033274.3085111

[60] Gali Noti and Vasilis Syrgkanis. 2021. Bid prediction in repeated auctions with learning. In Proceedings of the Web Conference 2021. 3953–3964. https://doi.org/10.1145/3442381.3449968

[61] Christos Papadimitriou and Georgios Piliouras. 2018. From Nash equilibria to chain recurrent sets: An algorithmic solution concept for game theory. Entropy 20, 10 (2018), 782.

[62] Julia Robinson. 1951. An iterative method of solving a game. Annals of mathematics 53, 1 (1951), 196–301.

[63] Tim Roughgarden. 2012. The price of anarchy in games of incomplete information. In Proceedings of the 13th ACM Conference on Electronic Commerce. 862–879.

[64] Tim Roughgarden, Vasilis Syrgkanis, and Eva Tardos. 2017. The price of anarchy in auctions. Journal of Artificial Intelligence Research 59 (2017), 59–101.

[65] Tim Roughgarden and Joshua R Wang. 2019. Minimizing regret with multiple reserves. ACM Transactions on Economics and Computation (TEAC) 7, 3 (2019).

[66] N. Shalev-Shwartz et al. 2011. Online learning and online convex optimization. Foundations and trends in Machine Learning 4 (2011), 107–194.

[67] Vasilis Syrgkanis and Eva Tardos. 2013. Composable and Efficient Mechanisms. In Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing (Palo Alto, California, USA) (STOC ’13). Association for Computing Machinery, New York, NY, USA, 211–220. https://doi.org/10.1145/2488608.2488635

[68] Moshe Tennenholtz and Oren Kurland. 2019. Rethinking search engines and recommendation systems: a game theoretic perspective. Commun. ACM 62, 12 (2019), 66–75.

[69] Hal R Varian. 2007. Position auctions. international Journal of industrial Organiza- tion 25, 6 (2007), 1163–1178.

[70] William Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. The Journal of finance 16, 1 (1961), 8–37.

[71] H Peyton Young. 2004. Strategic learning and its limits. OUP Oxford.

[72] Martin Zinkevich. 2003. Online convex programming and generalized infinitesimal gradient ascent. In Proceedings of the 20th international conference on machine learning. 928–936.
APPENDIX

A ADDITIONAL SIMULATIONS

Our simulations show similar convergence results for different variants of multiplicative-weights and FTPL algorithms. The following figures show the bid dynamics in second-price and first-price auctions when the first agent has a value of $v = 1$ and the second agent has a value of 0.5, for the Hedge algorithm and FTPL. We also look at simulations with a variant of FTPL with a recency bias that uses a geometric discount on the (perturbed) history of play. In all these different dynamics, in the second-price auction the high player always wins the auction and pays a price that is, on average, well below the actual second price of 0.5, and in the first-price auction the high player always wins and pays the second price.

A.1 Second-Price Auctions

![Bid dynamics in a repeated second-price auction with different types of regret-minimizing agents, where agent 1 has value 1 and agent 2 has value 0.5. Figure 7a: The exponential update version of multiplicative weights (the Hedge algorithm). Figure 7b: FTPL with a recency bias. Figure 7c: FTPL. The solid lines show running-window averages of the bids of each player over a window of 100 auctions in Figure 7a, and of 500 auctions in Figures 7b and 7c. The average price paid by the high bidder in sequences of 100 auctions in Figures 7a, and of 500 auctions in Figures 7b and 7c.](image)

Figure 7: Bid dynamics in a repeated second-price auction with different types of regret-minimizing agents, where agent 1 has value 1 and agent 2 has value 0.5. Figure 7a: The exponential update version of multiplicative weights (the Hedge algorithm). Figure 7b: FTPL with a recency bias. Figure 7c: FTPL. The solid lines show running-window averages of the bids of each player over a window of 100 auctions in Figure 7a, and of 500 auctions in Figures 7b and 7c. The average price paid by the high bidder in sequences of 100 auctions in Figures 7a, and of 500 auctions in Figures 7b and 7c. The average price paid by the high bidder in sequences of 100 auctions in Figures 7a, and of 500 auctions in Figures 7b and 7c. The average price paid by the high bidder in sequences of 100 auctions in Figures 7a, and of 500 auctions in Figures 7b and 7c.

A.2 First-Price Auctions

![Bid dynamics in a repeated first-price auction with different types of regret-minimizing agents, where agent 1 has value 1 and agent 2 has value 0.5. Figure 8a: The exponential update version of multiplicative weights (the Hedge algorithm). Figure 8b: FTPL with a recency bias. Figure 8c: FTPL. The solid lines show running-window averages of the bids of each player over a window of 100 auctions in Figure 8a, and of 500 auctions in Figures 8b and 8c.](image)

Figure 8: Bid dynamics in a repeated first-price auction with different types of regret-minimizing agents, where agent 1 has value 1 and agent 2 has value 0.5. Figure 8a: The exponential update version of multiplicative weights (the Hedge algorithm). Figure 8b: FTPL with a recency bias. Figure 8c: FTPL. The solid lines show running-window averages of the bids of each player over a window of 100 auctions in Figure 8a, and of 500 auctions in Figures 8b and 8c.

B CO-UNDOMINATED COARSE EQUILIBRIA

Proof. (Lemma 1): Consider any finite two-player game played repeatedly by two mean-based regret-minimizing agents, and assume that the dynamics converge to a distribution $p \in \Delta$. First, as is well known, since the agents are minimizing regret, $p$ must be a CCE (otherwise some agent must remain with positive average regret in the limit). Next, assume by way of contradiction that action $i$ of player 1 (w.l.o.g.) is weakly dominated by action $i'$ of player 1 in the support of player 2, and that $i$ is in the support of player 1. Let us use $u(i, j)$ to denote the utility for player 1 when player 1 bids $i$ and player 2 bids $j$, and $u_t(i)$ to denote the utility for player 1 from using the bid $i$ at time $t$, given the actual (empirical) action of player 2 at time $t$. Thus, for every action $j$ of player 2 with $P_{b \rightarrow j}(b_2 = j) > 0$ it holds that $u_t(i, j) \leq u_t(i', j)$, and there exists an action $j_0$ of player 2 with $P_{b \rightarrow j_0}(b_2 = j_0) = q_0 > 0$ such that $u_t(i, j_0) < u_t(i', j_0)$, and $P_{b \rightarrow j_0}(b_1 = i) = p_0 > 0$.

Denote $\delta = u(i', j_0) - u(i, j_0)$, and $\Delta_T = \sum_{t=1}^T u_t(i') - u_t(i)$. Let $\epsilon > 0$ s.t. $\epsilon < \frac{q_0 \delta}{\sqrt T}$ and $\delta < \frac{2 \delta}{\sqrt T}$, and take $T$ s.t. the algorithm used by player 1 is $\gamma(T)$-mean-based, and $T > T_0(\epsilon)$, and $\gamma(T) < \epsilon^2$.

Claim 1: Fix any $0 < t \leq T$. If $|p_i^T - p| \leq \epsilon$ it holds that $\Delta_T^U > \epsilon$. Proof: Notice that if $p_i^T = p$, then $\Delta_T^U = q_0 \delta t - \epsilon t$. If $p_i^T > p$, then $T_T^U > q_0 \delta - 2\epsilon \geq \frac{2 \delta^2}{q_0^2} > \epsilon$. On the other hand, from below, let us consider the case where $|p_i^T - p| = 2\epsilon$ and all the difference between the two distribution vectors in concentrated on action profiles in which the dominating action $i'$ has zero utility and the dominated action $i$ has a utility of 1. Thus $\Delta_T^U > (1 - \epsilon)q_0 \delta t - \epsilon t$, and therefore $\Delta_T^U > q_0 \delta - 2\epsilon \geq \frac{2 \delta^2}{q_0^2} > \epsilon$.

Claim 2: Denote $p_i^T(i) = \sum_j p_j^T(i, j)$. For every $t > T_1 = \sqrt \epsilon \cdot T$ with probability more than $1 - 2e$, it holds that $p_i^T(i) < 2\sqrt \epsilon$. Proof: By claim 1 above, for every $t > eT$, in the event that $|p_i^T - p| \leq \epsilon$ it holds that $\Delta_T^U > \epsilon t \geq \epsilon^2 T$, and thus from the mean-based property of player 1, it follows that the probability that action $i$ is chosen at time $t + 1$ is less than $\gamma(T) < \epsilon^2$. From the convergence assumption, this holds with probability at least $1 - \epsilon$ for any $t > eT$. Thus, with high probability (at least $1 - \epsilon$) the empirical average number of times that action $i$ is chosen in rounds $eT, ..., t$ is not far from its expectation of $e^2$, which is less than $\epsilon$. To bound $p_i^T(i)$ from above, consider the “worst case” where $p_{1+i}^T(i) = 1$ (until time $eT$). Thus, for $t > T_1$, with probability more than $1 - 2e$ it holds that $p_i^T(i) \leq \frac{1}{2}(eT + (t - eT)e) < \frac{1}{2}(eT + et) \leq \frac{1}{2}(1 + T) \epsilon = \epsilon + \sqrt \epsilon < 2\sqrt \epsilon$.

We thus have by claim 2 that with probability more than $1 - 2e$, $p_i^T(i) < 2\sqrt \epsilon$, and since $\delta < 2\sqrt \epsilon$, and $p_0 - 2\sqrt \epsilon > 3\sqrt \epsilon - 2\sqrt \epsilon = \sqrt \epsilon > \epsilon$, we have that $Pr\left(|p_i^T(i) - p| > \epsilon\right) > 1 - 2e$, in contradiction to the convergence assumption. Therefore, the dominated action $i$ cannot be in the support of player 1. This holds for the other player as well, and thus the distribution $\mathbf{p}$ must be co-undominated. □

Remark: Notice that this argument can be extended also to the case where the dynamics converge to $\mathbf{p}$ with a strictly positive probability that may be less than 1.
C SECOND-PRICE AUCTIONS

Proof. (Lemma 2): Denote the high player with value \( v \) as player 1 and the low player with value \( w \) as player 2. Assume that the players are in a CCE in which there is positive probability that player 1 loses the auction. Thus, there exist \( i_0, j_0 \) such that \( 0 = u_1(i_0, j_0) < u_2(w + \epsilon, j_0) = v - j_0 \). If for all \( j \in B \) (\( B \) being the support of player 2) it holds that \( u_1(i_0, j) \leq u_1(w + \epsilon, j) \), then \( w + \epsilon \) weakly dominates \( i_0 \), leading to a contradiction with the co-undominated CCE condition. If we assume the contrary, we have that there exists \( j \in B \) such that \( u_1(i_0, j) > u_1(w + \epsilon, j) \). This is impossible since bidding \( w + \epsilon \) wins the auction and gives utility \( v - j \), while any other bid gives either the same utility or zero. Therefore, in any co-undominated CCE the high player always wins.

D FIRST-PRICE AUCTIONS

To prove Lemma 3, we start by analyzing the symmetric case of bidders with equal values, which is technically slightly different.

We now proceed to prove Lemma 3:

Proof. (Lemma 3): Denote the value of player 1 as \( v \) and the value of player 2 as \( w \) and assume w.l.o.g. that \( v \geq w \). Lemma 4 deals with the symmetric case \( v = w + 2 \epsilon \). Here we will show that when \( v > w > 2 \epsilon \), bids that are not in \( [w - 2 \epsilon, \ldots, w + \epsilon] \) cannot be in the support of player 1.

First, for any support \( B \) of player 2, any bid \( b > w + \epsilon \) is strictly dominated for player 1 by bidding \( w + \epsilon \). Thus, in any co-undominated CCE, the high player does not bid above \( w + \epsilon \).

Next, let \( b_1, b_2 \) be the lowest bids of the players in the support and let \( B_1, B_2 \) be the highest bids in the support, and assume that \( b_1 < w - 2 \epsilon \). We will consider the following cases:

Case I: \( b_1 = B_1 \). Here player 1 plays a pure strategy and so the CCE condition implies that player 2 must best-reply to it. Thus, for \( b_1 = B_1 < w - 2 \epsilon \), we must have \( b_2 = B_2 = b_1 + \epsilon \), which is not an equilibrium, since player 1 would then always lose and thus prefers using a fixed bid \( w - 2 \epsilon \). Therefore, it must be that \( b_1 < B_1 \).

Case II: \( b_1 = b_2 \). Here, if \( b_1 = b_2 < w - 2 \epsilon \) then \( b_1 \) is strictly dominated by \( b_1 + \epsilon \) (b1 loses unless player 2 bids \( b_2 \), in which case player 1’s utility is \((v - b_1)/2 \) due to the tie, but then \( v - (b_1 + \epsilon), which this player gets from bidding \( b_1 + \epsilon \), is better). Thus, we have \( b_1 = b_2 \geq w - 2 \epsilon \).

Case III: \( b_1 < b_2 \). In this case \( b_1 \) always loses and always gets a utility of 0. Thus, due to the co-undominated CCE condition, player 1 must get zero utility for every bid in his support and every bid in player 2’s support, and so he gets zero utility. This leads to a contradiction since player 1 can bid \( w \) and have positive utility in every auction.

Case IV: \( b_1 > b_2 \). Here \( b_2 \) always loses and always gets zero utility. Thus, due to the co-undominated CCE condition, player 2 must get zero utility for every bid in the support. However, player 2 can bid \( w - 2 \epsilon \) and have a positive utility of \( 2 \epsilon \) in every auction in which player 1 bids \( b_1 \), a contradiction.

Hence, the support of the high player must be in \([w - 2 \epsilon, \ldots, w + \epsilon]\).

Next, let us show that the high player (with value \( v \)) wins the auction with probability approaching 1 when the grid step \( \epsilon \) is small compared to the differences in values. Assume that there is positive probability that player 1 bids \( a \) and player 2 bids \( b \) and \( a < b \). We know that \( w - 2 \epsilon \leq b \leq w \) and \( w - 2 \epsilon \leq a \leq w + \epsilon \), and denote \( \delta = v - w > 2 \epsilon \). On the one hand, the expected utility for player 1 from bidding \( a \) in a sequence of \( T \) auctions is at most \( T \cdot (1 - p)(v - w + \epsilon) \). On the other hand, the utility for player 1 from placing a fixed bid of \( w + \epsilon \), which always wins the auction, is \( T \cdot (v - w) \). Thus, from the CCE condition it follows that \( p \leq \frac{2 \epsilon}{v - w + \epsilon} \), and therefore \( p \) approaches zero as \( \epsilon / \delta \) approaches zero, i.e., as the grid becomes small compared to the difference in values.

E GENERALIZED FIRST-PRICE AUCTIONS

Here we provide the proof of Theorem 3 on the Nash equilibrium of the GFP auction. The full formal details of our analysis of GFP auctions can be found in the full version of the paper [46].

Proof. (Theorem 3): Denote the bid density function of player 1 in a Nash equilibrium by \( f(x) \) and the bid density function of player 2 in a Nash equilibrium by \( g(y) \), and denote the respective cumulative density functions by \( F(x) \) and \( G(y) \). First, observe that every bid \( y > w/2 \) of the second player is dominated by \( y = 0 \). Therefore, the support of the distributions \( f(x) \) and \( g(y) \) is \( [0, w/2] \). Next, the expected utility to player 1 from bid \( x \), assuming that player 2 plays according to \( g(y) \), is \( u_1(x) = \int_0^w g(y)(v - x)dy + \int_{w/2}^{w} g(y)\frac{v}{2}(v - x)dy = \frac{1}{2}(v - x)(G(x) + 1) \), where the first term is the expected payoff from winning the first slot in the auction, and the second term is the expected payoff from getting the second slot. The expected utility to player 2 from bid \( y \), assuming that player 1 plays according to \( f(x) \), is \( u_2(y) = \int_0^y f(x)(w - y)dx + \int_{y}^{w} f(x)\frac{w}{2}(w - y)dx = \frac{1}{2}(w - y)(F(y) + 1) \).

In equilibrium the derivatives of the utilities must satisfy

\[ \frac{dxu_1(x)}{dy} = (v - x)(G(x) - G(x) - 1) = 0, \quad \frac{dxu_2(y)}{dy} = (w - y)(F(y) + 1) - 1 = 0. \]

With another derivative, we obtain differential equations for \( f \) and \( g \). We also replace the variable names by their original notations:

\[ (v - y)(g(y) - 2g(y)) = 0, \quad (w - x)(f(x) - 2f(x)) = 0. \]

The solution to these equations is of the form \( f(x) = \frac{c_1}{(w - x)^{c_2}} \), \( g(y) = \frac{c_2}{(w - y)^{c_1}} \), where \( c_1, c_2 \) are constants that depend on the boundary conditions. The normalization constraint yields the unique solution: \( f(x) = \frac{w}{w - x} \), \( g(y) = (2 \frac{w}{w - y} - 1) \frac{w}{w - y} \). The integrals of these density functions lead to the cumulative densities stated in the theorem.

The utilities to the two players, with values \( v \) to player 1 and \( w \) to player 2 such that \( v \geq w \), in the Nash equilibrium are then given by the following integrals: \( u_1^{NE} = \int_0^w f(x)\int_0^{(1/2)(v - x)(G(x) + 1)}dx = \frac{w}{2} + (v - w)(1 - \ln(2)) \cdot u_1^{NE} = \int_0^w g(y)\int_0^{(1/2)(v - y)(F(y) + 1)}dy = \frac{w}{2} \). □

As \( y = w/2 \) is only weakly dominated by \( y = 0 \), the bid \( y = w/2 \) may have positive probability in equilibrium. In principle, if \( g(y) \) in equilibrium had a finite (atomic) weight on the bid \( y = w/2 \), then the support of \( f(x) \) would need to be extended to \([0, w/2 + \epsilon]\) for some \( \epsilon > 0 \). However, our solution for \( g(y) \) shows that the bid \( y = w/2 \) has zero probability and thus the support \([0, w/2]\) is sufficient for \( f(x) \).