Ex-Post Regret Learning in Games with Fixed and Random Matching: The Case of Private Values*

Rene Saran†

Maastricht University, Maastricht, The Netherlands

Roberto Serrano‡

Brown University, Providence, U.S.A.

IMDEA Social Sciences Institute, Madrid, Spain

This version: June 2010

Abstract

In contexts in which players have no priors, we analyze a learning process based on ex-post regret as a guide to understand how to play games of incomplete information under private values. The conclusions depend on whether players interact within a fixed set (fixed matching) or they are randomly matched to play the game (random matching). The relevant long run predictions are minimal sets that are closed under "the same or better reply" operations. Under additional assumptions in each case, the prediction boils down to pure Nash equilibria, pure ex-post equilibria or pure minimax regret equilibria. These three paradigms exhibit nice robustness properties in the sense that they are independent of beliefs about the exogenous uncertainty of type spaces. The results are illustrated in second-price auctions, first-price auctions and Bertrand duopolies.

Keywords: Fixed and Random Matching; Incomplete Information; Ex-Post Regret Learning; Nash Equilibrium; Ex-Post Equilibrium; Minimax Regret Equilibrium; Second-Price Auction; First-Price Auction; Bertrand Duopoly.

JEL: C72; C73; D43; D44; D82; D83.

1

1 Introduction

The standard analysis of games with incomplete information relies on the notion of Bayesian equilibrium, and has led to important theoretical constructions. Moreover, equilibrium insights can sometimes be applied with success to real-world problems. However, the use of equilibrium also has important limitations, as it assumes the existence of a common-knowledge type space to describe the underlying uncertainty. When for instance one talks to game theorists that have provided advice on how to bid in real auctions, this is often a major stumbling block: we may calculate an equilibrium of the given auction, but in the absence of common-knowledge of strategies, type spaces and prior beliefs, it is implausible to expect the equilibrium to be played. How should one bid then? More generally, how does a player behave in a game of incomplete information? Our analysis provides several possible answers, which should be treated as a guide to behavior in our benchmarks.

In agreement with recent trends in the robust analysis of game theory with incomplete information, we shall deemphasize the role of beliefs and turn to ex-post considerations. In fact, we shall propose an alternative paradigm to equilibrium theory by going all the way to the other extreme: without specifying any prior distribution, we shall analyze a learning model of behavior based on ex-post regrets. We apply our ex-post regret learning process to games in which each player knows his or her own payoff function. This makes our results applicable to private-values incomplete-information games. The analysis will be performed both under fixed matching and random matching of the players. We assume large multiple populations –one for each type of each player–, from which players are drawn to play the game.

Under fixed matching, a player is selected from each population and that set of players stays matched to play the game infinitely often. In the simplest version of the ex-post regret learning Markov process, each player starts by choosing his or her first action arbitrarily. After that first period, his or her last action is repeated with positive probability (inertia), but the player also switches to other actions with positive probability if and only if he or she regrets not having used them in the last period. That is, the player compares the payoff obtained in the last period and wonders what would have happened, *ceteris paribus*, had he or she used a different action.[1] The player feels regret for not having used those actions that would have

---

[1] See Saran and Serrano (2010) for an extension of the analysis to regret matching with respect to any finite memory, as opposed to only one-period memory, studied here.

strictly increased his or her last period's payoff, and switches to playing them with arbitrary positive probability. This means that, at any given period, a player's action is either unchanged or modified into one that constitutes a strictly better reply to the last action profile.

In general finite games in normal form, we show that the recurrent classes of our finite Markov process correspond to minimal sets of actions that are closed under "the same or better reply" (minimal CUSOBR sets). These sets are related to the minimal product sets that are "closed under weakly better replies" of Ritzberger and Weibull (1995). Any of the Ritzberger-Weibull closed sets contains one CUSOBR set, sometimes properly, since the "(weakly) better-reply" correspondence contains the "same or (strictly) better-reply" correspondence. However, we find that some minimal CUSOBR sets are not necessarily product sets, and sometimes they need not be included in any closed set à la Ritzberger and Weibull.[2]

For weakly acyclic games under better replies, the result is stronger. All the recurrent classes of the ex-post regret learning process are singletons, and the set of absorbing states coincides with the set of pure Nash equilibria of the complete-information game corresponding to the private valuations of the players that have been matched. Thus, in spite of the initial incomplete information, the set of players, through the experience developed by repeatedly playing the game together, arrive at Nash equilibrium play. The process converges with probability 1 to one of the pure Nash equilibria; which of them is selected will depend on the initial condition. Throughout the paper we use three running examples to illustrate our results, the second-price auction, the first-price auction and Bertrand duopoly competition. We show that all three are weakly acyclic games under better replies.

Under random matching, an independent draw from each player's population takes place every period. The selected players play for one period and then return to the pool. We conceive a game of incomplete information associated with the random matching model. A strategy for a player in this game is a function specifying for each type of that payer, the distribution of actions amongst the corresponding population. For each ex-post realization of the types, each player's payoff is a function of the strategies used in the associated populations. For a given player, two strategies are adjacent if exactly one type of that player has switched his or her action, while all other types have not, in the two strategies in question. The set of states of this Markov process is now the set of strategy profiles.

_____

[2]Sets that are closed under weakly better replies are curb (closed under rational behavior) sets. CUSOBR sets are not, as some better replies are excluded.

The recurrent classes of the ex-post regret learning process with random matching coincide with those minimal sets of strategy profiles that are "closed under the same or adjacent ex-post better replies" (minimal CUSOAEBR sets) of the incomplete information game described. For any strategy profile, the "same or adjacent ex-post better reply" correspondence consists of the same strategy profile –in which no player has changed his or her action– and those adjacent strategy profiles that are obtained when in some ex-post event, some types of the players switch to (strictly) better replies.

Once again, for games that are weakly acyclic with respect to adjacent ex-post better replies, the result is stronger. All recurrent classes are singletons and each absorbing state of the dynamics corresponds to a pure ex-post equilibrium of the incomplete information game. Some games, like the second-price auction, are weakly acyclic in this sense, but most games are not. Essentially the condition of weak acyclicity amounts to the existence of ex-post equilibrium, quite demanding in general games.

Moreover, in general games, the minimal CUSOAEBR sets can be rather large. We illustrate this by exhibiting the unique minimal CUSOAEBR sets for the first-price auction and Bertrand duopoly game. Almost every strategy profile is part of such a minimal set, which is therefore uninformative in terms of an economic prediction. In an attempt to gain determinacy, we turn to stochastic monotone dynamics by allowing players to "make mistakes" with small probability, and where the probability of switching to an action is an increasing function of the associated regret. Taking limits as mistakes probability vanishes would not help, however, since the problem is the large recurrent class of the unperturbed process, as just described.

Instead, we propose an approach in which taking another limit is possible and yields an interesting answer. We take the limit as the switch probabilities vanish.[3] If one performs such a limiting exercise, one obtains a selection of the "recurrent classes" of the associated "unperturbed process," which happens to be a snapshot of the game in which the behavior of all agents is fixed, and hence the set of "recurrent classes" simply coincides with the set of states of the process. Thus, the exercise performed in this part of the paper is the long-run prediction in a random matching context in which the behavioral rule is the perturbed monotone ex-post regret learning, but where switches are unlikely events. Such an application of stochastic stability provides an extremely powerful refinement in these cases.[4] For the first-price

---

[3]These results are robust to small non-negligible probabilities, as confirmed via simulations.

[4]We also introduce a test to check for stochastic stability, of interest in its own right, based on

auction and Bertrand duopoly, a unique strategy profile is selected. In the first-price auction, each bidder bids half his valuation. In Bertrand duopoly, each firm uses that price that, if it were the only firm in the market, would yield half the monopoly profit. In both cases, these actions are such that the maximum gain from increasing the bid/price equates the maximum gain from decreasing it. It follows that the strategy profiles selected by stochastic stability are minimax regret equilibria (Hyafil and Boutilier (2004)).[5] Our learning foundation of minimax regret equilibria in the first-price auction and Bertrand duopoly game is a consequence of the inverse of the regrets being the exponents of the switch probabilities, which go to zero in the stochastic stability analysis. We view this as a valid economic prediction in the absence of information about the exact distribution of types in the population; Savage (1951) was the first to propose minimizing the maximum ex-post regret – minimax regret – as a decision-making criterion in problems in which the agent has no *specific* prior about the probabilities of the states of the world.

For each version of the model, our study yields the relevant sets that are closed under the "same or better reply" as long-run predictions. Under stronger assumptions, play ends up at pure Nash equilibria of the corresponding complete information game (fixed matching in weakly acyclic games), pure ex-post equilibria (random matching in weakly acyclic games) and minimax regret equilibria (random matching and unlikely switches when there is no ex-post equilibrium). With respect to the exogenous uncertainty stemming from type spaces, these three paradigms are belief-independent and provide answers to how the game could be played, which ought to be compared to its Bayesian equilibria.[6]

1.1 Related Literature

Our processes are part of the no-regret learning literature (see Fudenberg and Levine (1998), Hart (2005) and Sandholm (2009) for different surveys of the area; see also Saran and Serrano (2010) and the references therein). However, our emphasis in the current paper concerns games of incomplete information. As just stated, Sav-

---

the construction of certain weighted cycles. This is especially useful in systems like ours, where the construction of minimal cost rooted trees would require to know the transitions between *any* pair of states.

[5]At a minimax regret equilibrium, each player uses a strategy that minimizes his or her maximum ex-post regret. Therefore, ex-post equilibria are always minimax regret equilibria (because players have no regrets), but the latter set is generally non-empty.

[6]For the reasons outlined above, more research on dynamics and learning processes applied to Bayesian games would be desirable (see Dekel, Fudenberg and Levine (2004) and Ely and Sandholm (2005) for early contributions).

age (1951) was the seminal contribution suggesting minimax regret as a decision rule in contexts in which agents have no prior beliefs. Noting that Bayesian-Nash equilibrium does not offer any guidance to behavior when the players lack common knowledge of priors, Linhart and Radner (1989) study minimax-regret strategies in $\frac{1}{2}$-double auctions. Hyafil and Boutilier (2004) define the minimax-regret equilibrium for such environments.

Especially related to our fixed-matching model is Hon-Snir, Monderer and Sela (1998), who model learning in an auction context differently. The bidders' valuations are determined first. Then, this fixed set of players repeatedly plays a first-price auction. Players are informed about the profile of bids at the end of each auction. The paper looks at two learning rules: generalized fictitious play and adaptive learning with bounded recall. The main result is that, if all players use either learning rule, then per-period play converges to Nash equilibrium of the one-shot auction in which players' types are common knowledge.[7]

Our last results use stochastic stability. Most applications of this methodology in non-cooperative game theory have been confined to games with complete information.[8] An exception is Jensen, Sloth and Whitta-Jacobsen (2005), which extends the perturbed best-response model in Young (1993) to finite Bayesian games.[9] Compared to our analysis, their players possess much more information, both ex-ante and ex-post.

1.2 Plan of the Paper

Section 2 describes the ex-post regret learning rule. Section 3 studies it under fixed matching, and Section 4 under random matching. Each of these two sections contains a subsection on weakly acyclic games that shows the stronger results. Section 4 contains also a subsection that deals with stochastic stability analysis. The ap-

---

[7]The particular equilibrium is the one in which the player with the highest valuation wins and pays a price equal to the second-highest valuation. However, this happens because of their *a priori* assumption that no player bids more than her valuation. We would also get a similar result with this additional assumption.

[8]See Young (1998) for an account of different applications.

[9]They make three assumptions that are necessary for this extension. First, the players know the true distribution of types in the population; second, the types of the matched players are truthfully revealed to everyone at the end of the interaction; and third, for each type of each player, there is a record of the action taken by that type during some past periods in which that type was selected. Under these assumptions, the unperturbed best-response dynamics, appropriately redefined, converges with probability one to some convention, which is a state that is "equivalent" to a strict Bayesian equilibrium of the game – if the latter exists. The perturbations then select among the different strict Bayesian equilibria.

plications to second-price auctions, first-price auctions and Bertrand duopolies are brought up to illustrate the relevant results in each case. Section 5 concludes, and Section 6 collects the proofs of the major results.

2   Ex-Post Regret Learning

Consider a N-person game of incomplete information G, with a finite set of types V_i and a finite set of actions A_i for each player who plays in position $i \in N$ (we use N for both the set and number of positions). There is a large population of players $\mathcal{P}_i$ for each position $i \in N$ (we will also use i to denote a player who plays in position i). We assume private values, and for the most part we use the word "valuation" (or "cost" depending on the application) instead of type. For each $i \in N$ and each $v_i \in V_i$, there exists a nonempty subset of players $\mathcal{P}_{v_i} \subseteq \mathcal{P}_i$ whose valuations equal v_i. A player's valuation is her private information, it is drawn once and for all, and remains constant over time. A player's type here involves no beliefs. The question we tackle is how the players behave under this severe informational restriction.

Players' interactions will take place under two distinct matching assumptions: fixed matching and random matching:

- **Fixed Matching.** For each position, a single player is selected randomly and independently from the corresponding population of players. The selected players are matched once and for all, and they repeatedly play the game every period.

- **Random Matching.** In each period, a new set of players is selected randomly and independently, one for each position from the corresponding population of players, and matched to play the game. After they have played the game, the matched players return to their respective populations. The process restarts in the next period with a new selection and matching.

Each player in population $\mathcal{P}_i$ is identified by her valuation v_i and her unique action a_i that she plays in the event she is matched. We shall assume that players adapt their actions from one period to the next using the following ex-post regret learning rule: suppose in period t a player with valuation v_i is choosing action a_i.

- If this player is not matched in period t, then she does not change her action in period $t+1$.

- If this player is matched in period t, then let $\pi_i(v_i, a_i, a_{-i})$ be her ex-post payoff in period t's game, where a_{-i} is the profile of actions of all other players in that game. Pick any $a_i' \in A_i \setminus \{a_i\}$. Had she chosen a_i' instead of a_i in that game, *ceteris paribus*, her payoff would have been $\pi_i(v_i, a_i', a_{-i})$.[10]

  Define $\Delta_i(v_i, a_i, a_i', a_{-i}) = \pi_i(v_i, a_i', a_{-i}) - \pi_i(v_i, a_i, a_{-i})$.

  If $\Delta_i(v_i, a_i, a_i', a_{-i}) > 0$, then we refer to this number as valuation v_i's *ex-post regret* from using a_i instead of a_i' against a_{-i}. In this case, letting $q(\cdot) \in (0,1)$ as a basic switch probability function, we assume that the player changes her action to a_i' in period $t+1$ with probability $q(\Delta_i(v_i, a_i, a_i', a_{-i})) > 0$ if and only if $\Delta_i(v_i, a_i, a_i', a_{-i}) > 0$, and she does not switch her action with a positive probability, i.e., $\sum_{a_i'} q(\Delta_i(v_i, a_i, a_i', a_{-i})) < 1$.

  All these events entailing switches in actions are independent across players and time.

This learning dynamics satisfies two properties: (i) inertia, i.e., a matched player does not switch her action with a positive probability, and (ii) a matched player of valuation v_i switches to a_i' in period $t+1$ with a positive probability whenever her ex-post regret from not using that action is positive. Several different specifications of $q(\cdot)$ fulfill these properties;[11] in our analysis, we of course fix $q(\cdot)$ to be one such function.

Our concern in the next sections is the identification of the long-run behavior under fixed and random matching of the players who follow the ex-post regret learning rule. The dynamic process under fixed matching $[\mathcal{M}_{FM/G}(q)]$ is defined over states that are profiles of actions taken by the fixed set of players. The dynamic process under random matching $[\mathcal{M}_{RM/G}(q)]$ is defined over states that specify the distributions of actions in the different populations of players. Both are finite Markov processes, and they are generally not irreducible, but they are still aperiodic, and convergence always obtains to one of their recurrent classes.[12] We shall therefore state our results in terms of the recurrent classes of the dynamics $\mathcal{M}_{FM/G}(q)$ and

---

[10]In our applications, if the bid/price of a_i' is tied at the highest bid/lowest price, then $\pi_i(v_i, a_i', a_{-i})$ is the expected payoff defined by the tie-breaking lottery specified in the rules of the game.

[11]For instance, let Δ^* be the maximum ex-post regret across all valuation types of all players and A^* be the maximum number of actions that any player has. Then any $q(\cdot)$ such that $q(x) \in \left[0, \frac{1}{A^*\Delta^*}\right)$ and $q(x) > 0 \iff x > 0$ will fulfill these properties. We provide another set of specifications in our stochastic stability analysis.

[12]A recurrent class is a set of states such that if the process reaches one of them, it will never leave the set, and such that it does not admit a proper subset of states with the same property. An absorbing state is a singleton recurrent class.

$\mathcal{M}_{RM/G}(q)$. When switches to actions happen after an agent has ex-post regret, the analysis will reveal which actions are taken in the long run by a fixed set of players (fixed matching) and which distributions of actions are more likely to emerge in the long run in the large populations (random matching).

3   Long-Run Behavior under Fixed Matching

Let $G(v_1, \ldots, v_N)$ denote the complete information game in which the valuations of the players are $\{v_1, \ldots, v_N\}$, the sets of actions are $(A_i)_{i \in N}$ and the payoffs are given by the functions $(\pi_i(v_i, \cdot, \cdot))_{i \in N}$. This will be the underlying game once a fixed matching happens, though recall that our players do not know the other players' valuations.

Let $A = \prod_{i \in N} A_i$. For any $(a_i, a_{-i}) \in A$, the *set of same-or-better replies* for the player of valuation v_i is

$$R_{v_i}(a_i, a_{-i}) = \{a_i' \in A_i | \text{either } a_i' = a_i \text{ or } \pi_i(v_i, a_i', a_{-i}) > \pi_i(v_i, a_i, a_{-i})\}.$$

Let $R_{G(v_1, \ldots, v_N)} : A \to A$ be the *same-or-better-reply correspondence* of the game $G(v_1, \ldots, v_N)$, i.e.,

$$R_{G(v_1, \ldots, v_N)}(a_1, \ldots, a_N) = \prod_{i \in N} R_{v_i}(a_i, a_{-i}).$$

Definition 3.1. A set of action profiles $\hat{A} \subseteq A$ in $G(v_1, \ldots, v_N)$ is *closed under same-or-better replies* (henceforth, CUSOBR set) if for all $(a_1, \ldots, a_N) \in \hat{A}$, we have $R_{G(v_1, \ldots, v_N)}(a_1, \ldots, a_N) \subseteq \hat{A}$. A *minimal CUSOBR set* is a CUSOBR set that does not contain a proper subset that is a CUSOBR set.

For any nonempty $\hat{A} \subseteq A$, define

$$\tilde{R}_{G(v_1, \ldots, v_N)}(\hat{A}) = \bigcup_{(a_1, \ldots, a_N) \in \hat{A}} \left(\prod_{i \in N} R_{v_i}(a_i, a_{-i}) \right).$$

Equivalently, $\hat{A}$ is a CUSOBR set if and only if $\hat{A}$ is a fixed point of $\tilde{R}_{G(v_1, \ldots, v_N)}$, i.e., $\tilde{R}_{G(v_1, \ldots, v_N)}(\hat{A}) = \hat{A}$.

It is easy to see that $(a_1, \ldots, a_N)$ is a pure Nash equilibrium of $G(v_1, \ldots, v_N)$ if and only if $\{(a_1, \ldots, a_N)\}$ is a singleton minimal CUSOBR set. Furthermore, since the game $G(v_1, \ldots, v_N)$ has a finite number of action profiles, there exists a minimal

CUSOBR set.

CUSOBR sets are related to product sets that are closed under weakly better replies (Ritzberger and Weibull (1995)); we discuss the differences at present. For any $(a_i, a_{-i}) \in A$, the *set of weakly-better replies* for the player of valuation v_i is

$$WBR_{v_i}(a_i, a_{-i}) = \{a_i' \in A_i | \pi_i(v_i, a_i', a_{-i}) \geq \pi_i(v_i, a_i, a_{-i})\}.$$

Let $WBR_{G(v_1,\ldots,v_N)} : A \to A$ be the *weakly-better-reply correspondence* of the game $G(v_1, \ldots, v_N)$, i.e.,

$$WBR_{G(v_1,\ldots,v_N)}(a_1, \ldots, a_N) = \prod_{i \in N} WBR_{v_i}(a_i, a_{-i}).$$

A *product set* of action profiles $\hat{A} \subseteq A$ is such that $\hat{A} = \prod_{i \in N} \hat{A}_i$, where $\emptyset \neq \hat{A}_i \subseteq A_i, \forall i \in N$. Then, $\hat{A}$ is a *product set of action profiles that is closed under weakly better replies* (or PCUWBR set) if $\hat{A}$ is a product set and $WBR_{G(v_1,\ldots,v_N)}(a_1, \ldots, a_N) \subseteq \hat{A}$ for all $(a_1, \ldots, a_N) \in \hat{A}$.[13] A *minimal PCUWBR set* is a PCUWBR set that does not contain a proper subset that is a PCUWBR set.

Since $R_{G(v_1,\ldots,v_N)}(a_1, \ldots, a_N) \subseteq WBR_{G(v_1,\ldots,v_N)}(a_1, \ldots, a_N)$, every PCUWBR set contains a CUSOBR set. Hence, every minimal PCUWBR set contains a minimal CUSOBR set. Therefore, in some games, the set of minimal CUSOBR sets is a refinement of the set of minimal PCUWBR sets. Game (a) in Figure 1 is an example; its unique minimal CUSOBR set $\{(D, R)\}$ is a refinement of its unique minimal PCUWBR set $\{(U, L), (U, R), (D, L), (D, R)\}$. However, it is not necessary that every minimal CUSOBR set of a game is a subset of some minimal PCUWBR set. Game (b) in Figure 1 is an example; $\{(D, L)\}$ is its unique minimal PCUWBR set whereas it has two minimal CUSOBR sets, $\{(D, L)\}$ and $\{(U, R)\}$. Finally, unlike minimal PCUWBR sets, minimal CUSOBR sets are not necessarily product sets. Game (c) in Figure 1 has a unique minimal CUSOBR set $\{(U, L), (M, L), (M, R), (D, L), (D, R)\}$, which is not a product set.

Consider any fixed matching in which $\{v_1, \ldots, v_N\}$ are the valuations of the se-

---

[13]This definition is equivalent to the original definition by Ritzberger and Weibull (1995) where the weakly-better-reply correspondence is defined over the domain of mixed strategies. For any $\hat{A}_i \subseteq A_i$, let $S(\hat{A}_i)$ be the set of mixed strategies with support in $\hat{A}_i$. The weakly-better-reply correspondence over the domain of mixed strategies, $\overline{WBR}_{G(v_1,\ldots,v_N)} : \prod_{i \in N} S(A_i) \to A$ is such that $\overline{WBR}_{G(v_1,\ldots,v_N)}(s_1, \ldots, s_N) = \prod_{i \in N} \overline{WBR}_{v_i}(s_i, s_{-i})$, where $\overline{WBR}_{v_i}(s_i, s_{-i}) = \{a_i' \in A_i | E_{s_{-i}}(\pi_i(v_i, a_i', a_{-i})) \geq E_{s_i, s_{-i}}(\pi_i(v_i, a_i, a_{-i}))\}, \forall i \in N$. The equivalence follows since for any product set $\hat{A} = \prod_{i \in N} \hat{A}_i$, $WBR_{G(v_1,\ldots,v_N)}(a_1, \ldots, a_N) \subseteq \hat{A}, \forall (a_1, \ldots, a_N) \in \hat{A} \iff \overline{WBR}_{G(v_1,\ldots,v_N)}(s_1, \ldots, s_N) \subseteq \hat{A}, \forall (s_1, \ldots, s_N) \in \prod_{i \in N} S(\hat{A}_i)$.

	L	R
U	1,3	3,1
D	2,0	3,2

(a)

	L	R
U	1,1	0,1
D	2,3	0,2

(b)

	L	R
U	1,0	0,0
M	0,1	2,0
D	2,0	0,1

(c)

Figure 1

lected players. Define a state of the matched players in a period as the current action profile $(a_1, \ldots, a_N)$ of the players. Hence, A is the set of states. Our interest in minimal CUSOBR sets stems from the following result:

Proposition 3.2. *Let $\{v_1, \ldots, v_N\}$ be the valuations of the matched players under fixed matching. Then, $\hat{A}$ is a recurrent class of $\mathcal{M}_{FM/G}(q)$ if and only if $\hat{A}$ is a minimal CUSOBR set of $G(v_1, \ldots, v_N)$.*

3.1 Weakly Acyclic Games

A stronger result can be established if $G(v_1, \ldots, v_N)$ is weakly acyclic under better replies. A *better-reply graph* is defined as follows: each action profile of $G(v_1, \ldots, v_N)$ is a vertex of the graph and there exists a directed edge from vertex $(a_1, \ldots, a_N)$ to vertex $(a'_1, \ldots, a'_N)$ if and only if $(a_1, \ldots, a_N) \neq (a'_1, \ldots, a'_N)$ and $(a'_1, \ldots, a'_N) \in R_{G(v_1, \ldots, v_N)}(a_1, \ldots, a_N)$. A *sink* is a vertex with no outgoing edges. A *better-reply path* is a sequence of vertices $(a_1^1, \ldots, a_N^1), \ldots, (a_1^L, \ldots, a_N^L)$ such that there exists a directed edge from each $(a_1^l, \ldots, a_N^l)$ to $(a_1^{l+1}, \ldots, a_N^{l+1})$. The game $G(v_1, \ldots, v_N)$ is *weakly acyclic under better replies* if from any action profile, there exists at least one better-reply path to a sink. Clearly, an action profile is a sink if and only if it is a pure Nash equilibrium of $G(v_1, \ldots, v_N)$. Thus, the game $G(v_1, \ldots, v_N)$ is weakly acyclic under better replies if from any action profile there exists at least one better-reply path to a pure Nash equilibrium.

Remark: Young (2004) defines a closely related class of weakly acyclic complete information games.[14] A *single-better-reply graph* is defined as follows: each action profile of $G(v_1, \ldots, v_N)$ is a vertex of the graph and there exists a directed edge from vertex $(a_1, \ldots, a_N)$ to vertex $(a'_1, \ldots, a'_N)$ if and only if there exists exactly one player such that $a'_i \neq a_i$, and $\pi_i(v_i, a'_i, a_{-i}) > \pi_i(v_i, a_i, a_{-i})$. A *single-better-reply path*

---

[14]Note that Young (2004) uses the term "weakly acyclic under better replies" to describe this class. We modify his terminology.

and a game being *weakly acyclic under single-better replies* are defined similarly as above. Again, an action profile is a sink if and only if it is a pure Nash equilibrium of $G(v_1, \ldots, v_N)$. Thus, the game $G(v_1, \ldots, v_N)$ is weakly acyclic under single-better replies if from any action profile there exists at least one single-better-reply path to a pure Nash equilibrium.

Since a single-better-reply graph is a subgraph of a better-reply graph, a game is weakly acyclic under single-better replies only if it is weakly acyclic under better replies. In fact, the former class of games is smaller than the latter. For example, consider the following game with three players (row, column and matrix) and two actions for each player:

|	L	R			L	R
U	$0,1,1$	$1,0,0$		U	$1,0,0$	$1,0,0$
D	$1,0,0$	$0,1,0$		D	$0,1,1$	$1,1,0$
	A				B	

Figure 2: A game that is weakly acyclic under better replies but not weakly acyclic under single-better replies.

The single-better-reply graph of this game has two sinks corresponding to the two pure Nash equilibria, (U, R, B) and (D, R, B), and the rest of action profiles form a connected component (see Figure 3). Thus, the game is not weakly acyclic under single-better replies. However, in addition to the directed edges in the single-better-reply graph, there exists one more directed edge from (D, L, A) to (D, R, B) in the better-reply graph (see Figure 4). Over this edge, two players switch to better replies: the column player from L to R and the matrix player from A to B. Thus, this game is weakly acyclic under better replies.

The result in the current subsection, stronger than the first proposition stated earlier, is this:

Lemma 3.3. *If $G(v_1, \ldots, v_N)$ is weakly acyclic under better replies, then $\hat{A}$ is a minimal CUSOBR set if and only if $\hat{A}$ is a singleton, i.e., $\hat{A} = \{(a_1, \ldots, a_N)\}$, and $(a_1, \ldots, a_N)$ is a pure Nash equilibrium of $G(v_1 \ldots, v_N)$.*

As a corollary to the above lemma and Proposition 3.2, we obtain the following result.
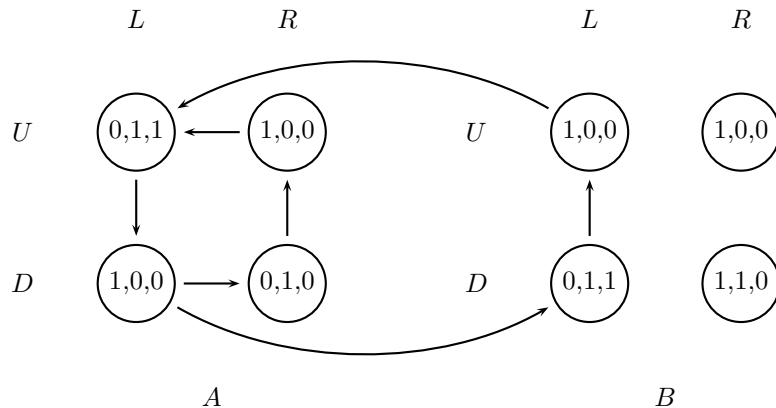
Figure 3: Single-Better-Reply Graph of the game in Figure 2
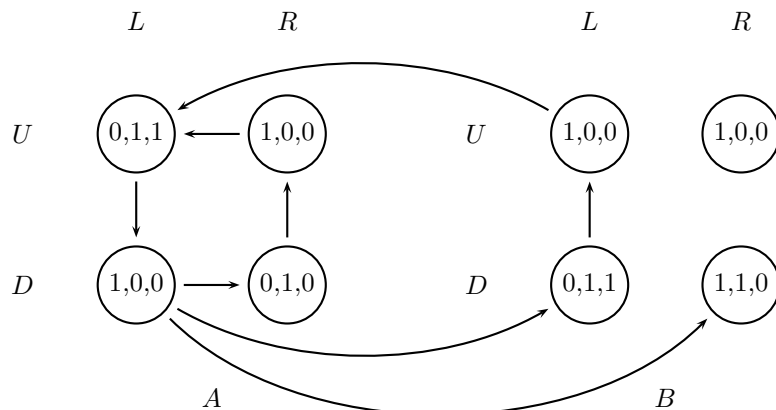
Figure 4: Better-Reply Graph of the game in Figure 2

Corollary 3.4. *Let $\{v_1, \ldots, v_N\}$ be the valuations of the matched players under fixed matching and suppose $G(v_1, \ldots, v_N)$ is weakly acyclic under better replies. All recurrent classes of $\mathcal{M}_{FM/G}(q)$ are singletons. Furthermore, $(a_1, \ldots, a_N)$ is an absorbing state if and only if it is a pure Nash equilibrium of $G(v_1, \ldots, v_N)$.*[15]

Remark: Thus, in spite of the initial incomplete information, ex-post regret learning under fixed matching gives players enough experience so that, when switches happen in the direction of ex-post regrets and the game is weakly acyclic under bet-

---

[15]Our dynamics is uncoupled in the sense of Hart and Mas-Colell (2003) and convergence to Nash equilibrium is obtained only in a specific class of games, while those authors seek a convergence result in *all* games. Our convergence to Nash equilibrium is a consequence of the fixed matching and the weakly acyclic property of the game.

ter replies, play in the game always converges to a pure-strategy Nash equilibrium of the complete information game corresponding to the valuations of the players in the match. Examples will be provided in our subsections on applications.

3.1.1 Applications: Second-Price Auctions, First-Price Auctions and Bertrand Duopolies

We now illustrate our results by studying specific games with our ex-post regret learning process under fixed matching. To begin, consider the second-price auction with $N \geq 2$ bidders and one seller. In each period, each seller is selling a single indivisible object and each bidder is interested in buying at most one unit of the good. Any seller of that indivisible object values it at 0 and this is known to the potential bidders. Let $\mathcal{S}_0$ denote the population of sellers. There are a large number of potential bidders, $\mathcal{P}_i$, who play in position $i \in N$. The valuation of any bidder in position i, $v_i \in Z \equiv \{0, \delta, 2\delta, \ldots, 1\}$.

Under fixed matching, a single seller is selected randomly from the population $\mathcal{S}_0$. The selected seller is then matched with $N \geq 2$ bidders, where bidder for position i is selected randomly and independently from $\mathcal{P}_i$. These matched players repeatedly play the second-price auction in every period. The rules of the auction are as follows: all the bidders matched to the seller simultaneously announce their bids for the object. A bidder can bid any number in Z. The bidder with the highest bid wins the object. The bidder who wins the object pays a price equal to the second-highest bid.[16] If the object is traded at price p, the winning bidder i's payoff is $v_i - p$ and all the other bidders matched to the seller get a payoff of 0. The seller's payoff is p. Each bidder in population $\mathcal{P}_i$ can be identified by her valuation v_i and her unique bid a_i that she bids in the event she is matched with a seller.

Let $\{v_1, \ldots, v_N\}$ be the valuations of the matched bidders under fixed matching. The set of states is Z^N, the collection of all possible bid profiles.

Lemma 3.5. *The one-shot second-price auction in which the valuations $\{v_1, \ldots, v_N\}$ of the N bidders are common knowledge is weakly acyclic under better replies.*

The following corollary follows from the above lemma and Corollary 3.4:

Corollary 3.6. *In the fixed-matching model, when the game is a second-price auction, all recurrent classes of $\mathcal{M}_{FM/SPA}(q)$ are singletons. Furthermore, $(a_1, \ldots, a_N)$*

---

[16]If more than one bidder bid the highest amount, then the object is allocated at random among the highest bidders and the price is equal to their bid.

is an absorbing state if and only if it is a pure Nash equilibrium of the one-shot second-price auction in which the valuations $\{v_1, \ldots, v_N\}$ of the N bidders are common knowledge.[17]

Now suppose that in every period, the seller uses the first-price auction to sell her object to the set of bidders that are matched to her. The only change in the rules is that the bidder who wins the object pays a price equal to her bid.

Lemma 3.7. *The one-shot first-price auction in which the valuations $\{v_1, \ldots, v_N\}$ of the N bidders are common knowledge is weakly acyclic under better replies.*

So we obtain the following corollary using the above lemma and Corollary 3.4:

Corollary 3.8. *In the fixed-matching model, when the game is a first-price auction, all recurrent classes of $\mathcal{M}_{FM/FPA}(q)$ are singletons. Furthermore, $(a_1, \ldots, a_N)$ is an absorbing state if and only if it is a pure Nash equilibrium of the one-shot first-price auction in which the valuations $\{v_1, \ldots, v_N\}$ of the N bidders are common knowledge.*

Consider next a game of Bertrand duopoly competition. The market demand for a good at price p is $Q = (x - p)/y$, where $x, y > 0$. There are a large number of potential sellers of the good. Let $\mathcal{P}_i$ be the population of sellers who play in position $i \in \{1, 2\}$. The cost of producing Q_i units of the good to a seller in position i is $c_i Q_i$, where $c_i \in C \equiv \{0, \delta, 2\delta, \ldots, x\}$ is her constant marginal cost (note that x is some multiple of δ). A seller's marginal cost is her private value, her type. For each $i \in \{1, 2\}$ and $c_i \in C$, there exists a nonempty set of sellers $\mathcal{P}_{c_i} \subseteq \mathcal{P}_i$ whose marginal costs equal c_i.

In Bertrand competition, the two matched sellers simultaneously post their prices p_1 and p_2 for the object, where each $p_i \in P = \{0, \frac{\delta}{2}, \delta, \ldots, x - \frac{\delta}{2}, x\}$. This ensures that each marginal cost type c_i can post her monopoly profit maximizing price $p^*(c_i) = \frac{x + c_i}{2}$. We assume that the buyers purchase the good from the cheaper seller and if both sellers charge the same price, then each seller captures the market, i.e., supplies the market demand, with equal probability. Hence, for any (p_1, p_2), seller i's ex-post payoff is 0 if either $p_i > p_j$ or $p_i = p_j$ and she looses the tie-breaking lottery, and

---

[17] Strictly speaking, applying Corollary 3.4 requires that the matched bidders are repeatedly playing a game. This is not the case here since the bidders adapt their bids in the ex-post stage. Therefore, in the event of a tie at the winning bid, the ex-post payoff of a bidder need not equal her expected payoff (where the expectation is with respect to the tie-breaking lottery). Nevertheless, given a_{-i}, the expected payoff of bidder of valuation v_i from bidding a'_i is greater than her expected payoff from bidding a_i if and only if in some ex-post event her regret $\Delta_i(v_i, a_i, a'_i, a_{-i}) > 0$. This fact and Lemma 3.5 are sufficient to establish the corollary. Similar comments apply to Corollaries 4.8, 3.8 and 3.10.

$\frac{(p_i-c_i)(x-p_i)}{y}$ if either $p_i < p_j$ or $p_i = p_j$ and she wins the tie-breaking lottery.[18] Each seller in population $\mathcal{P}_i$ can be identified by her marginal cost c_i and her unique price p_i that she posts in the event she is matched with another seller.

Under fixed matching, let $\{c_1, c_2\}$ be the marginal costs of the two matched sellers. A state of the matched sellers in a period is a profile of each seller's price, (p_1, p_2). Thus, the set of states is P^2.

Lemma 3.9. *The one-shot Bertrand duopoly game in which the marginal costs $\{c_1, c_2\}$ of the two sellers are common knowledge is weakly acyclic under better replies.*

The following corollary follows from the above lemma and Corollary 3.4:

Corollary 3.10. *In the fixed-matching model, when the game is Bertrand duopoly competition, all recurrent classes of $\mathcal{M}_{FM/BC}(q)$ are singletons. Furthermore, (p_1, p_2) is an absorbing state if and only if it is a pure Nash equilibrium of the one-shot Bertrand game in which the marginal costs $\{c_1, c_2\}$ of the two sellers are common knowledge.*

Remark: In all our applications, the ex-post regret learning dynamics converges to one of the pure Nash equilibria of the game; which equilibrium will be the limit of the dynamics is a function of the initial condition. The general point, nonetheless, is how Nash play emerges through using the learning process on the fixed set of players, in spite of the restrictive informational assumptions.

4 Long-Run Behavior under Random Matching

There is a game of incomplete information that corresponds to the random matching of the players. In this game, the set of strategies of valuation v_i, $\Sigma_i(v_i)$, can be identified with the set of distributions of actions A_i in the population $\mathcal{P}_{v_i}$. Let $\sigma_i(v_i)$ denote a strategy of valuation v_i. Given $\sigma_i(v_i)$, let $\sigma_i(v_i, a_i)$ be the relative frequency of action a_i and $A_i(\sigma_i(v_i))$ denote the support of $\sigma_i(v_i)$. The strategy of position i, σ_i, is a collection of strategies of all valuations $v_i \in V_i$. Let Σ_i be the set of strategies of position i. A strategy profile $(\sigma_1, \dots, \sigma_N)$ is a collection of strategies of all positions. For any $(\sigma_i, \sigma_{-i}) \in \Sigma$ and $(v_i, v_{-i}) \in V$, the expected payoff of position i is

$$\pi_i(v_i, \sigma_i(v_i), \sigma_{-i}(v_{-i})) = E_{\sigma_1(v_1),\dots,\sigma_N(v_N)}(\pi_i(v_i, a_i, a_{-i})).$$

---

[18]See Spulber (1995) for the equilibrium analysis of the winner-takes-all Bertrand competition with unknown but continuously distributed marginal costs.

Let $G(RM)$ be the N-position incomplete information game with valuations $(V_i)_{i \in N}$, strategies $(\Sigma_i)_{i \in N}$ and expected payoffs in each ex-post event $(v_1, \ldots, v_N)$ given by $\big(\pi_i(v_i, \sigma_i(v_i), \sigma_{-i}(v_{-i}))\big)_{i \in N}$.

For any $\sigma_i(v_i), \sigma_i'(v_i) \in \Sigma_i(v_i)$, let $d(\sigma_i(v_i), \sigma_i'(v_i))$ denote the Euclidean distance between $\sigma_i(v_i)$ and $\sigma_i'(v_i)$. We say that strategy σ_i' is *adjacent* to strategy σ_i if there exists a $v_i \in V_i$ such that $d(\sigma_i(v_i'), \sigma_i'(v_i')) = 0, \forall v_i' \neq v_i$, and $d(\sigma_i(v_i), \sigma_i'(v_i)) = \frac{\sqrt{2}}{|\mathcal{P}_{v_i}|}$, where $|\mathcal{P}_{v_i}|$ is the number of players in population $\mathcal{P}_{v_i}$ (i.e., in terms of the underlying distributions of actions in the population of players, σ_i' is adjacent to σ_i if exactly one player of valuation v_i plays differently in σ_i than in σ_i'). For any σ_i' that is adjacent to σ_i, let $v_i^{(\sigma_i, \sigma_i')} \in V_i$ be the unique valuation such that the distance between $\sigma_i(v_i^{(\sigma_i, \sigma_i')})$ and $\sigma_i'(v_i^{(\sigma_i, \sigma_i')})$ is positive. Furthermore, there exist exactly two actions in A_i, say $a_i^{(\sigma_i, \sigma_i')}$ and $a_i'^{(\sigma_i, \sigma_i')}$, such that

$$\sigma_i(v_i^{(\sigma_i, \sigma_i')}, a_i^{(\sigma_i, \sigma_i')}) = \sigma_i'(v_i^{(\sigma_i, \sigma_i')}, a_i^{(\sigma_i, \sigma_i')}) + \frac{1}{|\mathcal{P}_{v_i}|},$$

$$\text{and} \quad \sigma_i'(v_i^{(\sigma_i, \sigma_i')}, a_i'^{(\sigma_i, \sigma_i')}) = \sigma_i(v_i^{(\sigma_i, \sigma_i')}, a_i'^{(\sigma_i, \sigma_i')}) + \frac{1}{|\mathcal{P}_{v_i}|}.$$

For any $\sigma_i \in \Sigma_i$, let

$$\Sigma_i(\sigma_i) = \{\sigma_i' \in \Sigma_i | \text{either } \sigma_i' = \sigma_i \text{ or } \sigma_i' \text{ is adjacent to } \sigma_i\}.$$

Let $V = \prod_{i \in N} V_i$ and $\Sigma = \prod_{i \in N} \Sigma_i$. For any $(\sigma_1, \ldots, \sigma_N) \in \Sigma$, the action profiles that will be played with a positive probability conditional on the realization of types $(v_1, \ldots, v_N) \in V$ is given by $\prod_{i \in N} A_i(\sigma_i(v_i))$.

The *same-or-adjacent-ex-post-better-reply correspondence* of the game $G(RM)$, $R_{G(RM)} : \Sigma \to \Sigma$, is defined as follows: $(\sigma_1', \ldots, \sigma_N') \in R_{G(RM)}(\sigma_1, \ldots, \sigma_N)$ if and only if

1. $(\sigma_1', \ldots, \sigma_N') \in \prod_{i \in N} \Sigma_i(\sigma_i)$ and

2. if $(\sigma_1', \ldots, \sigma_N') \neq (\sigma_1, \ldots, \sigma_N)$, then let $I = \{i | \sigma_i \neq \sigma_i'\}$. There must exist $(v_1, \ldots, v_N) \in V$ and $(a_1, \ldots, a_N) \in \prod_{i \in N} A_i(\sigma_i(v_i))$ such that for all $i \in I$, we have $v_i = v_i^{(\sigma_i, \sigma_i')}$, $a_i = a_i^{(\sigma_i, \sigma_i')}$ and

$$\pi_i(v_i^{(\sigma_i, \sigma_i')}, a_i'^{(\sigma_i, \sigma_i')}, a_{-i}) > \pi_i(v_i^{(\sigma_i, \sigma_i')}, a_i^{(\sigma_i, \sigma_i')}, a_{-i}).$$

The first condition says that $(\sigma_1', \ldots, \sigma_N')$ is such that for all i, σ_i' is either the same or adjacent to σ_i. Now, if $(\sigma_1', \ldots, \sigma_N')$ differs from $(\sigma_1, \ldots, \sigma_N)$, then by the definition

of the set I, σ'_i is adjacent to σ_i for all $i \in I$. The second condition says that for all $i \in I$, this change in strategy should be justified as an ex-post better reply, i.e., there must exist an ex-post event in which for all $i \in I$, the realized valuation $v_i = v_i^{(\sigma_i, \sigma'_i)}$, the realized action $a_i = a_i^{(\sigma_i, \sigma'_i)}$, and the action $a_i'^{(\sigma_i, \sigma'_i)}$ is a better reply to a_{-i} than the action $a_i^{(\sigma_i, \sigma'_i)}$.

Definition 4.1. A set of strategy profiles $\hat{\Sigma} \subseteq \Sigma$ in $G(RM)$ is *closed under same-or-adjacent-ex-post-better replies* (henceforth, CUSOAEBR set) if for all $(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}$, we have $R_{G(RM)}(\sigma_1, \ldots, \sigma_N) \subseteq \hat{\Sigma}$. A *minimal CUSOAEBR set* is a CUSOAEBR set that does not contain a smaller CUSOAEBR set.

For any nonempty $\hat{\Sigma} \subseteq \Sigma$, define

$$\tilde{R}_{G(RM)}(\hat{\Sigma}) = \bigcup_{(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}} R_{G(RM)}(\sigma_1, \ldots, \sigma_N).$$

Equivalently, $\hat{\Sigma}$ is a CUSOAEBR set if and only if $\hat{\Sigma}$ is a fixed point of $\tilde{R}_{G(RM)}$, i.e., $\tilde{R}_{G(RM)}(\hat{\Sigma}) = \hat{\Sigma}$.

Since $G(RM)$ has a finite number of strategy profiles, there exists a minimal CUSOAEBR set.

Remark: Suppose there is a unique valuation type for all positions, i.e., $V_i = \{v_i\}, \forall i \in N$. Then $G(RM)$ is an extension of $G(v_1, \ldots, v_N)$ to mixed strategies Σ. Thus, $G(RM)$ is a complete information game in which the valuations of the players are $\{v_1, \ldots, v_N\}$, the sets of strategies are $(\Sigma_i)_{i \in N}$ and the payoffs are given by the functions $(\pi_i(v_i, \sigma_i, \sigma_{-i}))_{i \in N}$. In this case, if $\hat{\Sigma}$ is a CUSOAEBR set of $G(RM)$, then the set of action profiles that are in the support of $\hat{\Sigma}$, i.e.,

$$\bigcup_{(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}} \left(\prod_{i \in N} A_i(\sigma_i(v_i)) \right)$$

is a CUSOBR set of $G(v_1, \ldots, v_N)$.

Definition 4.2. A strategy profile $(\sigma_1, \ldots, \sigma_N) \in \Sigma$ is an *ex-post equilibrium* of $G(RM)$ if for all $i \in N$ and all $(v_i, v_{-i}) \in V$,

$$\pi_i(v_i, \sigma_i(v_i), \sigma_{-i}(v_{-i})) \geq \pi_i(v_i, \sigma'_i(v_i), \sigma_{-i}(v_{-i})), \forall \sigma'_i \in \Sigma_i.$$

Thus, in every ex-post state (v_i, v_{-i}), each player i's mixed strategy $\sigma_i(v_i)$ is a best response to the mixed-strategy profile $\sigma_{-i}(v_{-i})$.

A strategy profile $(\sigma_1, \ldots, \sigma_N) \in \Sigma$ is a *pure ex-post equilibrium* of $G(RM)$ if for all $i \in N$, $(v_i, v_{-i}) \in V$ and $(a_i, a_{-i}) \in \prod_{j \in N} A_j(\sigma_j(v_j))$,

$$\pi_i(v_i, a_i, a_{-i}) \geq \pi_i(v_i, a'_i, a_{-i}), \forall a'_i \in A_i.$$

There is a slight abuse of language in our use of the term "pure" here. Essentially, we are also allowing a mixed-strategy ex-post equilibrium (i.e., types can play non-degenerate mixed strategies) such that in every ex-post event, every action profile, which is in the support of the mixed-strategy profile corresponding to that ex-post event, is a pure Nash equilibrium. But there is a purification argument for any such mixed equilibrium, by having the appropriate proportions in the population of each type play the pure actions in the support of the mixed strategy.

Lemma 4.3. *A strategy profile $(\sigma_1, \ldots, \sigma_N)$ is a pure ex-post equilibrium of $G(RM)$ if and only if $\{(\sigma_1, \ldots, \sigma_N)\}$ is a singleton CUSOAEBR set.*

A state in a period is a list specifying for all $i \in N$ and $v_i \in V_i$, the distribution of actions A_i in population $\mathcal{P}_{v_i}$. Hence, a state is a strategy profile of $G(RM)$ and the set of states is Σ. Given G and random matching of the players, for fixed switch function $q(\cdot)$, consider now the learning rule presented in Section 2 and the associated Markov process $\mathcal{M}_{RM/G}(q)$ on the state space Σ.

Proposition 4.4. *$\hat{\Sigma}$ is a recurrent class of $\mathcal{M}_{RM/G}(q)$ if and only if $\hat{\Sigma}$ is a minimal CUSOAEBR set of $G(RM)$.*

4.1 Weakly Acyclic Games

As in the fixed-matching model, stronger results are obtained for weakly acyclic games. An *adjacent-ex-post-better-reply graph* is defined as follows: each strategy profile $(\sigma_1, \ldots, \sigma_N)$ of $G(RM)$ is a vertex of the graph. There exists a directed edge from vertex $(\sigma_1, \ldots, \sigma_N)$ to vertex $(\sigma'_1, \ldots, \sigma'_N)$ if and only if $(\sigma_1, \ldots, \sigma_N) \neq (\sigma'_1, \ldots, \sigma'_N)$ and $(\sigma'_1, \ldots, \sigma'_N) \in R_{G(RM)}(\sigma_1, \ldots, \sigma_N)$. An *adjacent-ex-post-better-reply path* is a sequence of vertices $(\sigma_1^1, \ldots, \sigma_N^1), \ldots, (\sigma_1^L, \ldots, \sigma_N^L)$ such that there exists a directed edge from each $(\sigma_1^l, \ldots, \sigma_N^l)$ to $(\sigma_1^{l+1}, \ldots, \sigma_N^{l+1})$. The game $G(RM)$ is *weakly acyclic under adjacent-ex-post-better replies* if from any strategy profile, there exists at least one adjacent-ex-post-better-reply path to a sink. Note that a strategy profile is a sink if and only if it is a pure ex-post equilibrium of $G(RM)$ (using Lemma 4.3). Thus, the game $G(RM)$ is weakly acyclic under adjacent-ex-post-better replies if

from any strategy profile there exists at least one adjacent-ex-post-better-reply path to a pure ex-post equilibrium.

Lemma 4.5. *If $G(RM)$ is weakly acyclic under adjacent-ex-post-better replies, then $\hat{\Sigma}$ is a minimal CUSOAEBR set if and only if $\hat{\Sigma}$ is a singleton, i.e., $\hat{\Sigma} = \{(\sigma_1, \ldots, \sigma_N)\}$, and $(\sigma_1, \ldots, \sigma_N)$ is a pure ex-post equilibrium of $G(RM)$.*

We obtain the following result as a corollary to the above lemma and Proposition 4.4.

Corollary 4.6. *Suppose $G(RM)$ is weakly acyclic under adjacent-ex-post-better replies. All recurrent classes of $\mathcal{M}_{RM/G}(q)$ are singletons. Furthermore, $(\sigma_1, \ldots, \sigma_N)$ is an absorbing state if and only if it is a pure ex-post equilibrium of $G(RM)$.*

Remark: Thus, when players are randomly matched and they use ex-post regret learning in a game that is weakly acyclic under adjacent-ex-post-better replies, all valuation types of all players will play in the long run according to some pure ex-post equilibrium of the game.

4.1.1 Application: Second-Price Auctions

Under random matching, in each period t, a single seller is selected randomly from the population $\mathcal{S}_0$. The selected seller is then matched with $N \geq 2$ bidders, where bidder for position i is selected randomly and independently from $\mathcal{P}_i$. These matched players play the second-price auction in period t. After the end of the auction, these players return to their respective populations. A new set of players is selected in period $t + 1$ and so on.

Let $SPA(RM)$ denote the incomplete information game corresponding to the random matching of the players when the matched players play the second-price auction. Thus, $SPA(RM)$ is the N-bidder one-shot second-price auction under incomplete information with valuations $V_i = Z, \forall i \in N$, and strategies $(\Sigma_i)_{i \in N}$, where Σ_i is the set of distributions of bids in populations $\mathcal{P}_{v_i}, \forall v_i \in Z$.

Lemma 4.7. *$SPA(RM)$ is weakly acyclic under adjacent-ex-post-better replies.*

The next corollary follows from the above lemma and Corollary 4.6:

Corollary 4.8. *In the random-matching model, when the game is a second-price auction, all recurrent classes of $\mathcal{M}_{RM/SPA}(q)$ are singletons. Furthermore, $(\sigma_1, \ldots, \sigma_N)$ is an absorbing state if and only if it is a pure ex-post equilibrium of $SPA(RM)$.*

Remark: One pure ex-post equilibrium is truth telling. However, there also exist pure ex-post equilibria of $SPA(RM)$ that are ex-post inefficient (this is also true for the standard one-shot second-price auction model in which the valuations are continuously distributed on a certain interval and bidders can use any mixed strategy). An example is the following strategy profile: let $n \in (0, 1/\delta)$ and

$$\sigma_1(v_1, n\delta) = 1, \forall v_1 \leq n\delta, \quad \text{and} \quad \sigma_1(v_1, v_1) = 1, \forall v_1 > n\delta,$$

and for all $i \neq 1$

$$\sigma_i(v_i, 0) = 1, \forall v_i \leq n\delta, \quad \text{and} \quad \sigma_i(v_i, v_i) = 1, \forall v_i > n\delta.$$

Hence, even such inefficient pure ex-post equilibria are absorbing states of the ex-post regret dynamics under random matching.[19]

4.2   Stochastic Stability

A necessary condition for a game to be weakly acyclic under adjacent-ex-post-better replies is that it has an ex-post equilibrium, which is often not satisfied (for instance, our applications to first-price auction and Bertrand competition in the sequel are not weakly acyclic under adjacent-ex-post-better replies). The recurrent classes of the process $\mathcal{M}_{RM/G}(q)$ will typically be very large, offering limited prediction.

In an attempt to provide a sharper prediction in these cases, we shall now specify $q(\Delta_i(v_i, a_i, a_i', a_{-i})) = q^{\frac{1}{\Delta_i(v_i, a_i, a_i', a_{-i})}}$, i.e., q is now a positive parameter that is sufficiently small to ensure positive probability of inertia. Thus, switches to other actions are an increasing function of the corresponding ex-post regret, and the way in which ex-post regret affects switching probabilities is exponential. But now, in addition, if $\Delta_i(v_i, a_i, a_i', a_{-i}) \leq 0$, then agent i changes her action to a_i' in period $t+1$ with probability $q^{\frac{1}{\gamma}}$, where $\gamma > 0$ is smaller than any ex-post regret.[20] Allowing for a positive γ leads to ergodic dynamics.

Letting $\alpha = q^{\frac{1}{\gamma}}$ be the probability of an individual's "mistake" –a switch under

---

[19]Note how if we had a single population model –which is a reasonable alternative model since the game is symmetric–, these inefficient equilibria would be eliminated. Consider the above strategy profile. With a single population, the bidders in position 1 will also meet each other, and hence, those who bid above their valuations will experience positive regret. This is the only result of our three applications where the assumption of single versus multiple populations would make any qualitative difference.

[20]This minimum (positive) ex-post regret is well-defined since the sets of players, valuations and actions are finite.

no regret–, we are now interested in analyzing the long run properties of the random process $\mathcal{M}_{RM/G}(q, \alpha)$ when α is a small but positive constant. This is an irreducible and aperiodic process, and it is a regular perturbation of $\mathcal{M}_{RM/G}(q, 0)$, which corresponds to the random matching ex-post regret learning process of Section 2. It is known that it admits a unique invariant distribution for each $\alpha > 0$ and $q > 0$. Further, the support of its limiting distribution as $\alpha \to 0$ (which is the same as $\gamma \to 0$) is contained within the recurrent classes of the corresponding Markov process with $\alpha = 0$. However, as we just observed, such recurrent classes are very large and therefore we propose to perform a different exercise.

We fix $\gamma > 0$ and take the limit with respect to $q \to 0$.[21] If one fixes $\gamma > 0$ and allows q to go to zero, the process $\mathcal{M}_{RM/G}(q, \alpha)$ is a regular perturbed process of $\mathcal{M}_{RM/G}(0, 0)$. Hence, as $q \to 0$, the support of the unique invariant distribution of $\mathcal{M}_{RM/G}(q, \alpha)$ is contained within the recurrent classes of $\mathcal{M}_{RM/G}(0, 0)$, which are all singletons (indeed, the set of absorbing classes of $\mathcal{M}_{RM/G}(0, 0)$ coincides with the set of states). Following Kandori, Mailath and Rob (1993) and Young (1993), we refer to the states that are in the support of this limiting distribution as *stochastically stable*. The limiting distribution approximates both the frequency with which a state is visited over a long horizon and the probability of being in a particular state at a point in time. Hence, the stochastically stable states are the only states on which the system will spend a positive proportion of time in the very long run when the switches of actions are possible but very unlikely events.[22]

Consider any two different states $\sigma = (\sigma_1, \ldots, \sigma_N)$ and $\sigma' = (\sigma'_1, \ldots, \sigma'_N)$ such that there is a positive probability of moving from the former to the latter in one period under the Markov process $\mathcal{M}_{RM/G}(q, \alpha)$. This probability is of the order q^r for some unique $r > 0$ as $q \to 0$. Define the *resistance* of going from σ and σ' as $r(\sigma, \sigma') = r$.[23]

A *weighted-adjacent-ex-post-better-reply graph* is an adjacent-ex-post-better-reply graph with weights assigned to all directed edges. Pick the directed edge from $\sigma = (\sigma_1, \ldots, \sigma_N)$ to $\sigma' = (\sigma'_1, \ldots, \sigma'_N)$ in the adjacent-ex-post-better-reply graph.

---

[21]The limit exercise allows us to obtain an analytical result. However, we point out that we have confirmed the result for fixed small values of q via simulations (available upon request).

[22]This analysis still respects the ordinality of payoffs in the game. Specifically, the results are robust to any transformation λ of the probability of switching as follows: a player with valuation v_i changes her action from a_i to $a'_i \in A_i \setminus \{a_i\}$ with probability $q^{\frac{1}{\lambda(\max\{\Delta_i(v_i, a_i, a'_i, a_{-i}), \gamma\})}}$, where $\lambda : \Re_{++} \to \Re_{++}$ is a strictly increasing function.

[23]To calculate $r(\sigma, \sigma')$, consider any match that can occur in state σ, which includes all the players who switch their action between σ and σ'. For each such match, sum the reciprocals of the ex-post regrets of the players who switch their action between σ and σ' (if the ex-post regret of any player is non-positive, then add $\frac{1}{\gamma}$ instead). The lowest such sum is $r(\sigma, \sigma')$.

Since $\sigma' \in R_{G(RM)}(\sigma)$, there is a positive probability of moving from σ to σ' in one period under the Markov Process $\mathcal{M}_{RM/G}(q, \alpha)$ (see the proof of Proposition 4.4). Define the *weight* of this directed edge as $r(\sigma, \sigma')$, which is positive by the definition of resistance. A *weighted-adjacent-ex-post-better-reply path* is a sequence of vertices $(\sigma_1^1, \ldots, \sigma_N^1), \ldots, (\sigma_1^L, \ldots, \sigma_N^L)$ and weights $w^1, \ldots, w^{L-1}$ such that there exists a directed edge from each $(\sigma_1^l, \ldots, \sigma_N^l)$ to $(\sigma_1^{l+1}, \ldots, \sigma_N^{l+1})$ with weight w^l. A *weighted cycle* is a weighted-adjacent-ex-post-better-reply path that begins and ends at the same vertex with no other repetition of vertices.

Next, we shall identify the stochastically stable states by eliminating states that pass the following weighted cycle test. This has interest in its own right as it provides a necessary condition for stochastic stability in processes like ours.

Proposition 4.9. *(Weighted cycle test)* $(\sigma_1, \ldots, \sigma_N)$ *is not stochastically stable if there exists an outgoing edge from* $(\sigma_1, \ldots, \sigma_N)$ *with weight less than* $\frac{1}{\gamma}$ *and either of the following hold:*

1. *There does not exist any weighted cycle containing this outgoing edge.*

2. *In all weighted cycles containing this outgoing edge, the weight of this outgoing edge is less than the weight of some other edge on the weighted cycle.*

This test substantially simplifies the search for the stochastically stable states. Since every state in Σ is absorbing in the process $\mathcal{M}_{RM/G}(0, 0)$, the standard technique of constructing rooted trees in order to find the minimal rooted trees would require calculating resistances between any two states in Σ. In contrast, in order to perform the weighted cycle test, we need to construct a single graph –the adjacent-ex-post-better-reply graph– and calculate the resistances of only the edges in this graph.

Remark: The necessary condition stated in the proposition is subtle. In particular, we cannot say that $(\sigma_1, \ldots, \sigma_N)$ is stochastically stable only if $(\sigma_1, \ldots, \sigma_N)$ belongs to a minimal CUSOAEBR set. For instance, let σ^1 be stochastically stable and not belong to any minimal CUSOAEBR set. We know that from any state σ^1 that does not belong to a minimal CUSOAEBR set, there exists a weighted-adjacent-ex-post-better-reply path $\sigma^1, \ldots, \sigma^L$ such that σ^L is in some minimal CUSOAEBR set. We also know that there does not exist any weighted cycle containing this *path*. However, there could exist a weighted cycle containing the first edge σ^1, σ^2.

4.2.1 Applications: First-Price Auctions and Bertrand Duopolies

Most games are not weakly acyclic under adjacent-ex-post-better-replies. As illustrations, consider first-price auctions and Bertrand duopolies. The analysis, through the use of stochastic stability, points in the direction of minimax-regret equilibrium (Hyafil and Boutilier (2004)). We remark that any ex-post equilibrium is also a minimax-regret equilibrium, but not vice versa. Indeed, for games with incomplete information that are not described by probabilistic beliefs (games with strict type uncertainty, in their terminology), these authors obtain a remarkable existence result for this concept.

Let $FPA(RM)$ denote the incomplete information game corresponding to the random matching of the players when the matched players play the first-price auction. Thus, $FPA(RM)$ is the N-bidder one-shot first-price auction under incomplete information with valuations $V_i = Z, \forall i \in N$, and strategies $(\Sigma_i)_{i \in N}$, where Σ_i is the set of distributions of bids in populations $\mathcal{P}_{v_i}, \forall v_i \in Z$.

Lemma 4.10. *Suppose $2\delta < 1$. $FPA(RM)$ has a unique minimal CUSOAEBR set $\hat{\Sigma}$. Furthermore, $\forall i \in N$, $v_i > 2\delta$, and $a_i \in [\delta, v_i - 2\delta] \bigcap Z$, there exists a $(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}$ such that $\sigma_i(v_i, a_i) = 1$.*

As a corollary of the above lemma and Proposition 4.4, it follows that when δ is small enough, the process $\mathcal{M}_{RM/FPA}(q)$ has a unique large recurrent class.[24] Hence, the answer obtained simply by looking at this recurrent class is not informative. As a possible way to increase predictive power, the next proposition identifies the stochastically stable states of $\mathcal{M}_{RM/FPA}(q, \alpha)$ as $q \to 0$.

Proposition 4.11. *In the random-matching model, when the game is a first-price auction, if $(\sigma_1, \ldots, \sigma_N)$ is stochastically stable in $\mathcal{M}_{RM/FPA}(q, \alpha)$ as $q \to 0$, then $\forall i \in N$ and $v_i \in Z$, we have $\sigma_i(v_i, a_i) > 0$ only if $a_i \in \left[\frac{v_i - \delta}{2}, \frac{v_i + \delta}{2}\right] \bigcap Z$.*

That is, when switches are unlikely events, all bidders bid approximately half their valuations almost all the time. We prove this by showing that a state that does not satisfy the condition in the proposition passes the weighted cycle test and hence, it is not stochastically stable. The intuition is as follows. Suppose in a state σ, the bidder i of valuation $v_i > \delta$ is bidding $z < v_i$ (the proof takes care of all cases). The bidder gains from adjusting her bid upwards only when it converts her from a looser to a winner, while she gains from adjusting her bid downwards only

---

[24]Applying Proposition 4.4 also requires that the randomly matched bidders are repeatedly playing a game; see footnote 17.

when it reduces the price of winning the object. If $z > \frac{v_i + \delta}{2}$, then the *maximum* gain from adjusting her bid downwards is equal to the highest price fall $z - \delta$ (when the highest opposing bid encountered by her is 0). Since all 0 valuation types must bid 0 – the unique no-regret bid for 0 valuation bidders– there is a positive probability that the last highest bid encountered by the bidder is 0 and thus, she could obtain the maximum gain of $z - \delta$ by adjusting her bid downwards. Let σ' be the new state after this bidder reduces her bid to δ, *ceteris paribus*. Since the bidder gains $z - \delta$ from this adjustment, the weight of the edge from σ to σ' in the weighted-adjacent-ex-post-better-reply graph is at most $\frac{1}{z-\delta} < \frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. There must be an edge in this cycle such that our bidder increases her bid from some $\tilde{z} < z$ to some $\hat{z} \geq z$. However, the *maximum* gain of the bidder in this transition is $v_i - \hat{z} \leq v_i - z < z - \delta$, when she turns from loosing at $\tilde{z}$ to winning at $\hat{z}$. Hence, the weight of this edge in the weighted cycle is greater than the weight of the edge from σ to σ'. Therefore, σ passes the weighted cycle test. On the other hand, if $z < \frac{v_i - \delta}{2}$, then the *maximum* gain from adjusting her bid upwards is at least $v_i - z^*$, where z^* is the lowest bid that is at least $\frac{v_i - \delta}{2}$. Now if there is another bidder j who bids z' such that $z \leq z' < \frac{v_i - \delta}{2}$ – existence is shown in the proof– then there is a positive probability that these two bidders meet and bidder i looses the auction (this happens when all other bidders in this match have 0 valuations). Then bidder i can obtain the gain of $v_i - z^*$ by adjusting her bid upwards to z^*. Let σ' be the new state after this bidder increases her bid to z^*, *ceteris paribus*. The weight of the edge from σ to σ' in the weighted-adjacent-ex-post-better-reply graph is at most $\frac{1}{v_i - z^*} < \frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. There must be an edge in this cycle such that our bidder changes her bid from z^* to some $\hat{z} \neq z^*$. If $\hat{z} > z^*$, then the *maximum* gain of the bidder in this transition is $v_i - \hat{z} < v_i - z^*$; whereas if $\hat{z} < z^*$, then the *maximum* gain of the bidder is $z^* - \delta < v_i - z^*$ (since $z^* < \frac{v_i + \delta}{2}$). Hence, the weight of this edge in the weighted cycle is greater than the weight of the edge from σ to σ'. Therefore, σ passes the weighted cycle test.

In fact, the stochastically stable bids (approximately) satisfy the following balance condition: at these bids, a bidder's maximum gain from increasing her bid (approx. $v_i - \sigma_i$) equals her maximum gain from decreasing it (approx. σ_i), leading to $\sigma_i = v_i/2$. Equivalently, at a stochastically stable bid, the maximum regret from not having used a higher bid is equal to the maximum regret from not having used a lower bid.

Remark: In the standard N-bidder one-shot first-price auction under incomplete information with the valuations for each bidder continuously distributed on $[0, 1]$, the

strategy profile in which each bidder bids half of her valuation is a minimax-regret equilibrium (Hyafil and Boutilier (2004)). Hence, Proposition 4.11 predicts that if the bidders repeatedly face a random first-price auction environment and they are very slow to adapt their bids, then in the long run, the distributions of bids in the populations of bidders will be a minimax-regret equilibrium.

Consider now Bertrand duopolies. Let $BC(RM)$ denote the incomplete information game corresponding to the random matching of the sellers when the matched sellers play the Bertrand duopoly competition game. Thus, $BC(RM)$ is the 2 player one-shot Bertrand competition game under incomplete information with marginal costs $V_i = C, \forall i \in \{1,2\}$, and strategies $(\Sigma_i)_{i \in \{1,2\}}$, where Σ_i is the set of distributions of prices in populations $\mathcal{P}_{c_i}, \forall c_i \in C$.

Lemma 4.12. $BC(RM)$ *has a unique minimal CUSOAEBR set* $\hat{\Sigma}$. *Furthermore,* $\forall i \in \{1,2\}$, $c_i \in C \setminus \{0, x\}$, *and* $p_i \in [c_i + \frac{\delta}{2}, p^*(c_i)] \bigcap P$, *there exists a* $(\sigma_1, \sigma_2) \in \hat{\Sigma}$ *such that* $\sigma_i(c_i, p_i) = 1$.

As a corollary of the above lemma and Proposition 4.4, it follows that the process $\mathcal{M}_{RM/BC}(q)$ has a unique large recurrent class.[25] Hence, the answer obtained simply by looking at this recurrent class is not informative. As in the first-price auction model, we identify the stochastically stable states of $\mathcal{M}_{RM/BC}(q, \alpha)$ as $q \to 0$.

If seller i with marginal cost c_i were a monopolist, then the price of $\hat{p}(c_i) = p^*(c_i) - \frac{x - c_i}{2\sqrt{2}}$ gives the seller a payoff equal to half of her maximum payoff –it is the unique such price that is less than $p^*(c_i)$. Clearly, there does not exist any price in P that equals $\hat{p}(c_i)$. Let $p^+(c_i)$ be the least element in P that is at least $\hat{p}(c_i)$.

Proposition 4.13. *In the random-matching model, when the game is Bertrand duopoly competition, if* (σ_1, σ_2) *is stochastically stable in* $\mathcal{M}_{RM/BC}(q, \alpha)$ *as* $q \to 0$, *then* $\forall i \in \{1,2\}$ *and* $c_i \in C$, *we have* $\sigma_i(c_i, p_i) > 0$ *only if* $p_i \in [p^+(c_i), p^+(c_i) + \delta] \bigcap P$.

Thus, when switches are unlikely events, any seller with marginal cost c_i will post approximately the price $p^+(c_i)$ almost all the time. This is because any state that does not satisfy the condition in the proposition passes the weighted cycle test and hence, it is not stochastically stable. The intuition is as follows. Suppose in a state σ, the seller i with marginal cost $c_i < x$ posts a price $p < p^+(c_i)$. Then $p < \hat{p}(c_i)$ and so the *maximum* gain from adjusting her price upwards is greater than half of her maximum monopoly payoff $\frac{(x - c_i)^2}{4y}$ (when the opponent's price is greater than

---

[25] Again, see footnote 17.

$p^*(c_i)$ and she switches from p to $p^*(c_i)$). Since all sellers with marginal costs equal to x must post the price x – the unique no-regret price for such sellers –, there is a positive probability that the last price encountered by the seller is x and thus, she could obtain the maximum gain by adjusting her price upwards to $p^*(c_i)$. Let σ' be the new state after this seller increases her price to $p^*(c_i)$, *ceteris paribus*. Since the seller gains an amount greater than $\frac{(x-c_i)^2}{8y}$ from this adjustment, the weight of the edge from σ to σ' in the weighted-adjacent-ex-post-better-reply graph is less than $\frac{8y}{(x-c_i)^2} < \frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. There must be an edge in this cycle such that our seller decreases her price from some $\tilde{p} > p$ to some $\bar{p} \leq p$. However, the *maximum* gain of the seller in this transition is less than $\frac{(x-c_i)^2}{8y}$ since $\bar{p} < \hat{p}(c_i)$. Hence, the weight of this edge in the weighted cycle is greater than the weight of the edge from σ to σ'. Therefore, σ passes the weighted cycle test. On the other hand, if $p > p^+(c_i) + \delta$, then the *maximum* gain from adjusting her price downwards is greater than $\frac{(x-c_i)^2}{8y}$ (e.g., when she switches from zero market share at price p to full market share at price $p^+(c_i)$). Now if there is a seller j who posts a price p_j such that $p \geq p_j > p^+(c_i)$ – existence is shown in the proof –, then there is a positive probability that these two sellers meet and seller i obtains a zero market share. Then seller i can obtain a gain greater than $\frac{(x-c_i)^2}{8y}$ by adjusting her price downwards to $p^+(c_i)$. Let σ' be the new state after this seller decreases her price to $p^+(c_i)$, *ceteris paribus*. The weight of the edge from σ to σ' in the weighted-adjacent-ex-post-better-reply graph is less than $\frac{8y}{(x-c_i)^2} < \frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. There must be an edge in this cycle such that our seller changes her price from $p^+(c_i)$ to some $\tilde{p} \neq p^+(c_i)$. Thus either $p > p^+(c_i)$ or $p < \hat{p}(c_i)$. Hence, irrespective of the value of p, the *maximum* gain of the seller is less than $\frac{(x-c_i)^2}{8y}$. Therefore, the weight of this edge in the weighted cycle is greater than the weight of the edge from σ to σ' and hence, σ passes the weighted cycle test.

Like in the first-price auction, the stochastically stable prices in Bertrand competition (approximately) satisfy a balance condition: at these prices, a seller's maximum gain from increasing her price equals her maximum gain from decreasing it.

Remark: In the one-shot Bertrand duopoly game under incomplete information with the marginal costs for each seller continuously distributed on $[0, x]$ (e.g., see Spulber (1995)), the strategy profile in which each seller posts the price equal to $\hat{p}(c_i)$ is a minimax-regret equilibrium. Hence, as in the first-price auction model, Proposition 4.13 predicts that if the sellers repeatedly face a random Bertrand competition environment and they are very slow to adapt their prices, then in the long

run, the distributions of prices in the populations of sellers will be a minimax-regret equilibrium.

5   Conclusion

We have analyzed a learning process based on ex-post regret as a guide to understand how to play games of incomplete information under private values. The conclusions depend on whether players interact within a fixed set (fixed matching) or they are randomly matched to play the game (random matching). The relevant long run predictions are minimal sets that are closed under "the same or better reply" operations. Under additional assumptions in each case, the prediction boils down to pure Nash equilibrium play, pure ex-post equilibrium play or pure minimax regret equilibrium play. These three paradigms exhibit nice robustness properties in the sense that they are independent of beliefs about the exogenous uncertainty of type spaces. At the very least, these predictions ought to be compared to the more standard one given by Bayesian equilibria. One obvious next step for further research would be to generalize these results for the case of interdependent values.

6   Proofs

Proof of Proposition 3.2: We first argue that $\hat{A}$ is a CUSOBR set if and only if whenever the dynamics $\mathcal{M}_{FM/G}(q)$ reaches any state in $\hat{A}$, it does not leave $\hat{A}$. Suppose $\hat{A}$ is such that whenever the dynamics reaches any state in $\hat{A}$, it never leaves $\hat{A}$. We argue that $R_{G(v_1,\ldots,v_N)}(a_1,\ldots,a_N) \subseteq \hat{A}$ for all $(a_1,\ldots,a_N) \in \hat{A}$. Pick any $(a_1',\ldots,a_N') \in R_{G(v_1,\ldots,v_N)}(a_1,\ldots,a_N)$. The argument is trivial if $(a_1',\ldots,a_N') = (a_1,\ldots,a_N)$. So suppose $(a_1',\ldots,a_N') \neq (a_1,\ldots,a_N)$. Let I be the set of players such that $a_i' \neq a_i$. Then for all $i \in I$, action a_i' is a better reply for valuation v_i to a_{-i} than a_i. There is a positive and independent probability that a player either does not switch her action or switches her action if she has positive regret. Hence, starting from state $(a_1,\ldots,a_N)$, there is a positive probability that, for all $i \in I$, the player of valuation v_i switches her action to a_i' while for all $j \notin I$, the player of valuation v_j does not switch her action. Hence, the dynamics moves from $(a_1,\ldots,a_N)$ to $(a_1',\ldots,a_N')$ in one period with a positive probability. Therefore, $(a_1',\ldots,a_N') \in \hat{A}$. Hence, $\hat{A}$ is a CUSOBR set. Next, suppose $\hat{A}$ is a CUSOBR set. Then for all $(a_1,\ldots,a_N) \in \hat{A}$, we have $R_{G(v_1,\ldots,v_N)}(a_1,\ldots,a_N) \subseteq \hat{A}$. Since a player switches her action with a positive probability only if she has positive regret, this means that once the dynamics reaches

28

a state in $\hat{A}$, it never leaves $\hat{A}$. It follows that $\hat{A}$ is a recurrent class if and only if it is a minimal CUSOBR set. $\qquad\square$

Proof of Lemma 3.3: If $(a_1, \ldots, a_N)$ is a pure Nash equilibrium of the game, then clearly $\{(a_1, \ldots, a_N)\}$ is a minimal CUSOBR set. Now, pick any $a^1 = (a_1^1, \ldots, a_N^1)$ that is not a pure Nash equilibrium of the game and suppose a^1 is an element of some minimal CUSOBR set $\hat{A}$. Since $G(v_1, \ldots, v_N)$ is weakly acyclic under better replies, there exists a pure Nash equilibrium a^L and a better-reply path $a^1, \ldots, a^L$ such that for any a^l, a^{l+1} in the path, $a^{l+1} \in R_{G(v_1, \ldots, v_N)}(a^l)$. Since $a^1 \in \hat{A}$, it must also be that $a^L \in \hat{A}$. But $\{a^L\}$ is a minimal CUSOBR set, a contradiction. $\qquad\square$

Proof of Lemma 3.5: Suppose $a^1 = (a_1^1, \ldots, a_N^1)$ is not a pure Nash equilibrium of the one-shot second-price auction in which the valuations $\{v_1, \ldots, v_N\}$ of the players are common knowledge. Then there exists bidder $i(1)$ who could gain by unilaterally deviating. Since bidding equal to one's valuation is a weakly dominant strategy, unilaterally deviating to bidding $v_{i(1)}$ is a better reply to $a_{-i(1)}^1$ for bidder $i(1)$. Let a^2 be the bid profile such that bidder $i(1)$ bids equal to her valuation and all other bidders $j \neq i(1)$ continue to bid $a_j^2 = a_j^1$. If a^2 is a pure Nash equilibrium of the game, then we are done. Otherwise, there exists a bidder $i(2) \neq i(1)$ who could gain by unilaterally deviating when the bid profile is a^2 (bidder $i(1)$ cannot gain a positive amount by unilateral deviation since she is playing her weakly dominant strategy). Again, have $i(2)$ bid truthfully and repeat the process. Since all bidders bidding truthfully is a pure Nash equilibrium of the game, it should be clear that in a finite number of steps, we will reach a bid profile a^L that is a pure Nash equilibrium of the game. Thus, the game is weakly acyclic under better replies. $\qquad\square$

Proof of Lemma 3.7: Suppose $a^1 = (a_1^1, \ldots, a_N^1)$ is not a pure Nash equilibrium of the one-shot first-price auction in which the valuations $\{v_1, \ldots, v_N\}$ are common knowledge. Then there exists a bidder $i(1)$ who could gain by unilaterally deviating. If bidder $i(1)$ is getting a negative expected payoff at a^1, let $a_{i(1)}^2 = 0$, which, with no other change to a^1, leads to a^2. If a^2 is a Nash equilibrium, we are done, and if not, first make the same change in the bids of those players who are receiving a negative expected payoff. Suppose that, after such changes, the bid profile is a^l. At a^l all bidders are receiving non-negative payoffs. If a^l is a Nash equilibrium, we are done. If not, bidder $i(l)$ has a profitable deviation, and there are only two possibilities: either she could gain by increasing her bid or decreasing her bid. To gain by increasing her

29

bid, she must weakly outbid the highest bid, while to gain from decreasing her bid, either she must reduce her loss or increase her profit.

First, suppose bidder $i(l)$ could gain by increasing her bid. Let a^{l*} be the highest bid in the bid profile a^l. Thus, bidder $i(l)$ can weakly outbid the highest bid a^{l*} and make a positive profit. Then, in particular, bidding either a^{l*} or $a^{l*}+\delta$ is a best reply to $a^l_{-i(l)}$ for bidder $i(l)$. Let a^{l+1} be the bid profile such that bidder $i(l)$ bids such a best reply to $a^l_{-i(l)}$, and all other bidders $j \neq i(l)$ continue to bid $a^{l+1}_j = a^l_j$. Note that the result of this is that the price in the auction is at least as high as before. If a^{l+1} is a pure Nash equilibrium of the game, then we are done. Otherwise, there exists a bidder $i(l+1) \neq i(l)$ who could gain by unilaterally deviating when the bid profile is a^{l+1}. This means that bidder $i(l+1)$ would like to weakly outbid the highest bid $a^{(l+1)*}$ in the bid profile a^{l+1}. Let $a^{l+2}_{i(l+1)}$ be a best reply to $a^{l+1}_{-i(l+1)}$, which leads to profile a^{l+2}. Note that, again, the price in the auction when the profile a^{l+2} is used is at least as high as the previous price. Again, either a^{l+2} is a Nash equilibrium and we are done, or it is not. But then, repeating the same argument a finite number of times, this process, based on a non-decreasing price sequence, must end at a Nash equilibrium.

Second, suppose bidder $i(l)$ can gain by decreasing her bid, and recall once again that all bidders are receiving non-negative expected payoffs. If at a^l, player $i(l)$ can profitably deviate by decreasing her bid, noting that her expected payoff is non-negative, she must be winning the auction but she can still win at a lower price. Thus, it must be that $a^{l*}_{-i(l)} < a^l_{i(l)} \leq v_{i(l)}$, where $a^{l*}_{-i(l)}$ is the highest bid among the bids $a^l_{-i(l)}$. Let $a^{l+1}_{i(l)} \geq a^{l*}_{-i(l)}$ be a best reply to $a^l_{-i(l)}$, thus defining a^{l+1}. Now, either a^{l+1} is a Nash equilibrium and we are done, or not. If not, it must be that one of the other bidders $(j \neq i(l))$ has a profitable deviation. If this deviation consists of increasing her bid, then we are back in the first case. So suppose that in a^{l+1}, no bidder has an improving deviation that consists of increasing her bid. Thus, for bidder $j \neq i(l)$, the improving deviation consists of decreasing her bid. Since bidder $i(l)$ bids the highest in a^{l+1}, it must be that $a^{l+1}_j = a^{l+1}_{i(l)}$ and bidder j obtains a negative expected payoff in a^{l+1}. For any such j, bidding $a^{l+2}_j = a^{l+1}_{i(l)} - \delta$ is a better reply than a^{l+1}_j. In this way, change the bid of any such bidder j to obtain the profile a^{l+2}. The highest (but not necessarily unique) bidder in a^{l+2} is $i(l)$ and she does not have an improving deviation in a^{l+2} (in particular, she cannot gain by reducing her bid to $a^{l+2}_{i(l)} - \delta$ because such a deviation at best ties her at the bid of $a^{l+2}_{i(l)} - \delta$). Clearly, none of the bidders who switched between a^{l+1} and a^{l+2} have an improving deviation. Finally, the rest of the bidders did not want to change their bid when the

highest opposing bid was $a_{i(l)}^{l+1} = a_{i(l)}^{l+2}$ and so, they also do not have an improving deviation. Hence, a^{l+2} is a Nash equilibrium $\qquad\square$

Proof of Lemma 3.9: This proof and subsequent ones for Bertrand duopoly are quite close to those for the first-price auctions, and hence we omit them (available upon request).

Proof of Lemma 4.3: It is easy to see that if $(\sigma_1, \ldots, \sigma_N)$ is a pure ex-post equilibrium, then it must be that $R_{G(RM)}(\sigma_1, \ldots, \sigma_N) = \{(\sigma_1, \ldots, \sigma_N)\}$. Now, suppose $(\sigma_1, \ldots, \sigma_N)$ is not a pure ex-post equilibrium. Then it must be that there exists a $i \in N$, $(v_i, v_{-i}) \in V$, $(a_i, a_{-i}) \in \prod_{j \in N} A_j(\sigma_j(v_j))$ and $a_i' \in A_i$ such that $\pi_i(v_i, a_i', a_{-i}) > \pi_i(v_i, a_i, a_{-i})$. Define σ_i' as follows: for all $v_i' \neq v_i$, let $\sigma_i'(v_i') = \sigma_i(v_i')$ and

$$
\sigma_i'(v_i, a_i'') = \begin{cases} \sigma_i(v_i, a_i'') + \frac{1}{|\mathcal{P}_{v_i}|}, & \text{if } a_i'' = a_i' \\ \sigma_i(v_i, a_i'') - \frac{1}{|\mathcal{P}_{v_i}|}, & \text{if } a_i'' = a_i \\ \sigma_i(v_i, a_i''), & \text{otherwise.} \end{cases}
$$

Then $(\sigma_i', \sigma_{-i}) \in R_{G(RM)}(\sigma_i, \sigma_{-i})$. $\qquad\square$

Proof of Proposition 4.4: We first argue that $\hat{\Sigma}$ is a CUSOAEBR set if and only if whenever the dynamics $\mathcal{M}_{RM/G}(q)$ reaches any state in $\hat{\Sigma}$, it does not leave $\hat{\Sigma}$. Suppose $\hat{\Sigma}$ is such that whenever the dynamics reaches any state in $\hat{\Sigma}$, it never leaves $\hat{\Sigma}$. We argue that $R_{G(RM)}(\sigma_1, \ldots, \sigma_N) \subseteq \hat{\Sigma}, \forall(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}$. Pick any $(\sigma_1', \ldots, \sigma_N') \in R_{G(RM)}(\sigma_1, \ldots, \sigma_N)$. If $(\sigma_1', \ldots, \sigma_N') = (\sigma_1, \ldots, \sigma_N)$, then the argument is trivial. So suppose $(\sigma_1', \ldots, \sigma_N') \neq (\sigma_1, \ldots, \sigma_N)$ and let I be the set of positions i such that $\sigma_i' \neq \sigma_i$. We know that $(\sigma_1', \ldots, \sigma_N') \in \prod_{i \in N} \Sigma_i(\sigma_i)$ and that there exist $(v_1, \ldots, v_N) \in V$ and $(a_1, \ldots, a_N) \in \prod_{i \in N} A_i(\sigma_i(v_i))$ such that for all $i \in I$, we have $v_i = v_i^{(\sigma_i, \sigma_i')}$, $a_i = a_i^{(\sigma_i, \sigma_i')}$ and

$$
\pi_i(v_i^{(\sigma_i, \sigma_i')}, a_i'^{(\sigma_i, \sigma_i')}, a_{-i}) > \pi_i(v_i^{(\sigma_i, \sigma_i')}, a_i^{(\sigma_i, \sigma_i')}, a_{-i}).
$$

In state $(\sigma_1, \ldots, \sigma_N)$, there is a positive probability that the valuation profile of the randomly matched players is $(v_1, \ldots, v_N)$ and these players play the action profile $(a_1, \ldots, a_N)$ since $(a_1, \ldots, a_N) \in \prod_{i \in N} A_i(\sigma_i(v_i))$. There is a positive and independent probability that a player either does not switch her action or switches her action if she has positive regret. Hence, there is a positive probability that, for all $i \in I$, the player of valuation v_i switches to $a_i'^{(\sigma_i, \sigma_i')}$ while for all $j \notin I$, the player of valua-

tion v_j continues to play a_j. Furthermore, there is no change in the distributions of actions in all other populations because none of the players from these populations is matched. Hence, there is a positive probability that for all $i \in I$, the distributions of actions in population $\mathcal{P}_i$ in the next period will be σ_i' while for all $j \notin I$, the distributions of actions in population $\mathcal{P}_j$ in the next period will remain σ_j. Thus, the dynamics will move from $(\sigma_1, \ldots, \sigma_N)$ to $(\sigma_1', \ldots, \sigma_N')$ in one period with a positive probability. Therefore, it must be that $(\sigma_1', \ldots, \sigma_N') \in \hat{\Sigma}$. Hence, $\hat{\Sigma}$ is a CUSOAEBR set. Next, suppose $\hat{\Sigma}$ is a CUSOAEBR set. Then for all $(\sigma_1, \ldots, \sigma_N) \in \hat{\Sigma}$, we have $R_{G(RM)}(\sigma_1, \ldots, \sigma_N) \subseteq \hat{\Sigma}$. A player switches her action with a positive probability only if she has positive regret and at most a single player in any population $\mathcal{P}_i$ switches her action in one period (since only one such player is matched). Hence, once the dynamics reaches a state in $\hat{\Sigma}$, it never leaves $\hat{\Sigma}$. It follows that $\hat{\Sigma}$ is a recurrent class if and only if it is a minimal CUSOAEBR set. $\square$

Proof of Lemma 4.5: If $(\sigma_1, \ldots, \sigma_N)$ is any pure ex-post equilibrium of the game, then clearly $\{(\sigma_1, \ldots, \sigma_N)\}$ is a minimal CUSOAEBR set. Now, pick any $\sigma^1 = (\sigma_1^1, \ldots, \sigma_N^1)$ that is not a pure ex-post equilibrium of the game and suppose σ^1 is an element of some minimal CUSOBR set $\hat{\Sigma}$. Since $G(RM)$ is weakly acyclic under adjacent-ex-post-better replies, there exists a pure ex-post equilibrium σ^L and an adjacent-ex-post-better-reply path $\sigma^1, \ldots, \sigma^L$ such that for any σ^l, σ^{l+1} in the path, $\sigma^{l+1} \in R_{G(RM)}(\sigma^l)$. Since $\sigma^1 \in \hat{\Sigma}$, it must also be that $\sigma^L \in \hat{\Sigma}$. But $\{\sigma^L\}$ is a minimal CUSOAEBR set, a contradiction. $\square$

Proof of Lemma 4.7: Suppose $\sigma^1 = (\sigma_1^1, \ldots, \sigma_N^1)$ is not a pure ex-post equilibrium of $SPA(RM)$. Then there exists a $i \in N$, an ex-post realization of types, (v_i, v_{-i}), and action profile, $(a_i, a_{-i}) \in \prod_{j \in N} A_j(\sigma_j^1(v_j))$, such that bidder i could gain by unilaterally deviating in this ex-post event. Since bidding equal to one's valuation is a weakly dominant strategy, it must be that $a_i \neq v_i$. Moreover, unilaterally deviating to bidding v_i is a better reply for bidder i in this ex-post event. Let σ^2 be the strategy profile such that $\sigma_j^2 = \sigma_j^1$ for all $j \neq i$ and σ_i^2 is adjacent to σ_i^1 with $\sigma_i^2(v_i, a_i) < \sigma_i^1(v_i, a_i)$ and $\sigma_i^2(v_i, v_i) > \sigma_i^1(v_i, v_i)$. If σ^2 is a pure ex-post equilibrium of the game, then we are done. Otherwise, repeat the above argument. Since the strategy profile in which all types of all bidders only bid truthfully is a pure ex-post equilibrium of the game, it should be clear that in a finite number of steps, we will reach a strategy profile σ^L that is a pure ex-post equilibrium of the game. Thus, the game is weakly acyclic under adjacent-ex-post-better replies. $\square$

Proof of Proposition 4.9: Before we prove this proposition, let's define the resistance $r(\sigma, \sigma')$ of going from state σ to state σ' when the probability of moving from σ to σ' in one period is zero under the Markov process $\mathcal{M}_{RM/G}(q, \alpha)$. Consider all paths $\sigma^1, \ldots, \sigma^L$ from $\sigma = \sigma^1$ to $\sigma' = \sigma^L$ such that there is a positive probability of moving from each σ^l to σ^{l+1} in one period. Then $r(\sigma, \sigma')$ is the lowest total resistance over all such paths, and the path with the lowest total resistance is called the *path of least resistance* from σ to σ'.

The following lemma will be used in the proof.

Lemma 6.1. *Suppose there is a positive probability of moving from σ to σ' in one period under the Markov process $\mathcal{M}_{RM/G}(q, \alpha)$. If $\sigma' \notin R_{G(RM)}(\sigma)$, then $r(\sigma, \sigma') \geq \frac{1}{\gamma}$.*

Proof: Let $\sigma = (\sigma_1, \ldots, \sigma_N)$ and $\sigma' = (\sigma'_1, \ldots, \sigma'_N)$. Since there is a positive probability of moving from σ to σ' in one period, it must be that $\sigma'_i \in \Sigma_i(\sigma_i), \forall i$ (there is a single match in every period). However, $\sigma' \notin R_{G(RM)}(\sigma)$. Then let $I = \{i | \sigma_i \neq \sigma'_i\}$. It follows that for all $(v_1, \ldots, v_N) \in V$ and $(a_1, \ldots, a_N) \in \prod_{i \in N} A_i(\sigma_i(v_i))$ with $v_i = v_i^{(\sigma_i, \sigma'_i)}$ and $a_i = a_i^{(\sigma_i, \sigma'_i)}$ for all $i \in I$, there exists a $j \in I$ such that

$$\pi_j(v_j^{(\sigma_j, \sigma'_j)}, a_j'^{(\sigma_j, \sigma'_j)}, a_{-j}) \leq \pi_j(v_j^{(\sigma_j, \sigma'_j)}, a_j^{(\sigma_j, \sigma'_j)}, a_{-j}).$$

Hence, in every possible match, there exists a player who switches her action without ex-post regret. Therefore, $r(\sigma, \sigma') \geq \frac{1}{\gamma}$. $\qquad\square$

Let σ, σ' be the outgoing edge described. By construction, $\sigma' \in R_{G(RM)}(\sigma)$. Consider an σ-rooted tree with the lowest total resistance. Add the directed edge σ, σ' to the σ-rooted tree. In the resulting graph, there exists a directed cycle $\sigma^1, \ldots, \sigma^L$ that starts and ends at σ. We argue that there exists a l with $2 \leq l < L$, such that $r(\sigma^l, \sigma^{l+1}) > r(\sigma, \sigma')$. For all l such that $2 \leq l < L$, let $\sigma[l]^1, \ldots, \sigma[l]^{K_l}$ be the path of least resistance from σ^l to σ^{l+1}.

- If there exists a l', with $2 \leq l' < L$, and k, with $1 \leq k < K_{l'}$, such that $\sigma[l']^{k+1} \notin R_{G(RM)}(\sigma[l']^k)$, then $r(\sigma[l']^k, \sigma[l']^{k+1}) \geq \frac{1}{\gamma}$ (using Lemma 6.1). Then, $r(\sigma^{l'}, \sigma^{l'+1}) \geq \frac{1}{\gamma} > r(\sigma, \sigma')$.

- Next suppose that for all l such that $2 \leq l < L$ and for all k such that $1 \leq k < K_l$, $\sigma[l]^{k+1} \in R_{G(RM)}(\sigma[l]^k)$. This, in addition to the fact that $\sigma' \in R_{G(RM)}(\sigma)$, implies that there exists a weighted cycle in the weighted-adjacent-ex-post-better-reply graph that contains the edge σ, σ'. Hence, if σ is such that there does not exist any weighted cycle containing the edge σ, σ',

then we have a contradiction and we are done. Otherwise, there exists an edge on this weighted cycle whose weight is greater than the weight of the edge σ, σ'. Hence, there must exist a l, with $2 \leq l < L$, such that $r(\sigma^l, \sigma^{l+1}) > r(\sigma, \sigma')$.

Now, delete the branch σ^l, σ^{l+1} from the directed cycle $\sigma^1, \ldots, \sigma^L$ to obtain an σ^l-rooted tree with a lower total resistance than the stochastic potential of σ. Thus, σ is not stochastically stable. $\qquad\square$

Proof of Lemma 4.10: Recall the definition of $\tilde{R}_{G(RM)}$ and accordingly define $\tilde{R}_{FPA(RM)}$. Let $\hat{\sigma} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_N)$ be such that $\hat{\sigma}_i(0,0) = \hat{\sigma}_i(\delta, 0) = 1$ and $\hat{\sigma}_i(v_i, \delta) = 1, \forall v_i > \delta$. Pick any $\sigma = (\sigma_1, \ldots, \sigma_N)$ and consider the iteration

$$\tilde{R}_{FPA(RM)}(\{\sigma\}) \subseteq \tilde{R}^2_{FPA(RM)}(\{\sigma\}) \subseteq \ldots \subseteq \tilde{R}^l_{FPA(RM)}(\{\sigma\}) \subseteq \ldots$$

We argue that there exists a $\hat{l}$ such that $\hat{\sigma} \in \tilde{R}^{\hat{l}}_{FPA(RM)}(\{\sigma\})$. Suppose σ is such that there exists a 0 valuation type of some player who plays a positive bid with a positive probability. Let z be the greatest positive bid played by any 0 valuation type of any player. Without loss of generality, let player i be such that $\sigma_i(0, z) > 0$. Let $\sigma^1_i \in \Sigma_i(\sigma_i)$ be such that $\sigma^1_i(0,0) > \sigma_i(0,0)$ and $\sigma^1_i(0, z) < \sigma_i(0, z)$. Then the strategy profile $\sigma^1 = (\sigma^1_1, \ldots, \sigma^1_N)$ such that $\sigma^1_j = \sigma_j, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\sigma\})$. This is because in the ex-post event $(v_1, \ldots, v_N)$ such that $v_j = 0, \forall j \in N$ and $(a_1, \ldots, a_N) \in \prod_{j \in N} A_j(\sigma_j(0))$ such that $a_i = z$, player i obtains a negative expected payoff and hence, the bid of 0 is a better reply to a_{-i}. If σ^1 is such that there exists a 0 valuation type of some player who plays a positive bid with a positive probability, then repeat this argument until we reach a σ^l such that all 0 valuation types of all players surely bid 0. By construction, $\sigma^l \in \tilde{R}^l_{FPA(RM)}(\{\sigma\})$. Suppose σ^l is such that there exists a δ valuation type of some player who plays a positive bid with a positive probability. Let z' be the greatest positive bid played by any δ valuation type of any player. Without loss of generality, let player i be such that $\sigma_i(\delta, z') > 0$. Let $\sigma^{l+1}_i \in \Sigma_i(\sigma^l_i)$ be such that $\sigma^{l+1}_i(\delta, 0) > \sigma^l_i(\delta, 0)$ and $\sigma^{l+1}_i(\delta, z') < \sigma^l_i(\delta, z')$. Then the strategy profile $\sigma^{l+1} = (\sigma^{l+1}_1, \ldots, \sigma^{l+1}_N)$ such that $\sigma^{l+1}_j = \sigma^l_j, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\sigma^l\})$ and hence, in $\tilde{R}^{l+1}_{FPA(RM)}(\{\sigma\})$. This is because in the ex-post event $(v_1, \ldots, v_N)$ such that $v_j = 0, \forall j \neq i$ and $v_i = \delta$, and $(a_1, \ldots, a_N) \in \prod_{j \in N} A_j(\sigma^l_j(v_j))$ such that $a_j = 0, \forall j \neq i$ and $a_i = z'$, player i obtains at most an expected payoff of 0 and hence, the bid of 0 is a better reply to a_{-i}. If σ^{l+1} is such that there exists a δ valuation type of some player who plays a positive bid with a positive probability, then repeat this argument until we reach a $\sigma^{l'}$ such that all 0 and δ valuation types

34

of all players surely bid 0. By construction, $\sigma^{l'} \in \tilde{R}^{l'}_{FPA(RM)}(\{\sigma\})$. Suppose $\sigma^{l'}$ is such that there exists a valuation type $v'_i > 2\delta$ of player i who bids $z'' \neq \delta$ with a positive probability. Let $\sigma^{l'+1}_i \in \Sigma_i(\sigma^{l'}_i)$ be such that $\sigma^{l'+1}_i(v'_i, \delta) > \sigma^{l'}_i(v'_i, \delta)$ and $\sigma^{l'+1}_i(v'_i, z'') < \sigma^{l'}_i(v'_i, z'')$. Then the strategy profile $\sigma^{l'+1} = (\sigma^{l'+1}_1, \dots, \sigma^{l'+1}_N)$ such that $\sigma^{l'+1}_j = \sigma^{l'}_j, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\sigma^{l'}\})$ and hence, in $\tilde{R}^{l'+1}_{FPA(RM)}(\{\sigma\})$. This is because in the ex-post event $(v_1, \dots, v_N)$ such that $v_j = 0, \forall j \neq i$ and $v_i = v'_i$, and $(a_1, \dots, a_N) \in \prod_{j \in N} A_j(\sigma^{l'}_j(v_j))$ such that $a_j = 0, \forall j \neq i$ and $a_i = z''$, the expected payoff of player i is either v'_i/N if $z'' = 0$ or $v'_i - z''$ if $z'' > \delta$ and hence, in any case, the bid of δ with the payoff $v'_i - \delta$ is a better reply to a_{-i}. If $\sigma^{l'+1}$ is such that there exists a valuation type $v'_j > 2\delta$ of some player j who plays a bid other than δ with a positive probability, then repeat this argument until we reach a $\sigma^{l''}$ such that all 0 and δ valuation types of all players surely bid 0 and all valuation types of all players with valuations greater than 2δ surely bid δ. By construction, $\sigma^{l''} \in \tilde{R}^{l''}_{FPA(RM)}(\{\sigma\})$. Suppose $\sigma^{l''}$ is such that there exists a valuation type 2δ of player i who plays a bid $\hat{z} \neq \delta$ with a positive probability. Let $\sigma^{l''+1}_i \in \Sigma_i(\sigma^{l''}_i)$ be such that $\sigma^{l''+1}_i(2\delta, \delta) > \sigma^{l''}_i(2\delta, \delta)$ and $\sigma^{l''+1}_i(2\delta, \hat{z}) < \sigma^{l''}_i(2\delta, \hat{z})$. Then the strategy profile $\sigma^{l''+1} = (\sigma^{l''+1}_1, \dots, \sigma^{l''+1}_N)$ such that $\sigma^{l''+1}_j = \sigma^{l''}_j, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\sigma^{l''}\})$ and hence, in $\tilde{R}^{l''+1}_{FPA(RM)}(\{\sigma\})$. This is because in the ex-post event $(v_1, \dots, v_N)$ such that $v_j = 3\delta, \forall j \neq i$ and $v_i = 2\delta$, and $(a_1, \dots, a_N) \in \prod_{j \in N} A_j(\sigma^{l''}_j(v_j))$ such that $a_j = \delta, \forall j \neq i$ and $a_i = \hat{z}$, the payoff of player i is at most 0 and hence, the bid of δ is a better reply to a_{-i}. If $\sigma^{l''+1}$ is such that there exists valuation type 2δ of some player j who plays a bid other than δ with a positive probability, then repeat this argument until we reach $\hat{\sigma}$. By construction, $\hat{\sigma} \in \tilde{R}^{\hat{l}}_{FPA(RM)}(\{\sigma\})$ for some $\hat{l}$.

It follows that any CUSOAEBR set must contain $\hat{\sigma}$ and hence, there exists a unique minimal CUSOAEBR set.

Pick any player i of valuation $v'_i \in [3\delta, 1]$ and $a_i \in [\delta, v'_i - 2\delta] \bigcap Z$. If $v'_i = 3\delta$, then $a_i = \delta$ and we already have $\hat{\sigma} \in \hat{\Sigma}$ such that $\hat{\sigma}_i(3\delta, \delta) = 1$. So pick $v'_i \in [4\delta, 1]$. Let $\hat{\sigma}^1_i \in \Sigma_i(\hat{\sigma}_i)$ be such that $\hat{\sigma}^1_i(v'_i, 2\delta) > \hat{\sigma}_i(v'_i, 2\delta)$ and $\hat{\sigma}^1_i(v'_i, \delta) < \hat{\sigma}_i(v'_i, \delta)$. Then the strategy profile $\hat{\sigma}^1 = (\hat{\sigma}^1_1, \dots, \hat{\sigma}^1_N)$ such that $\hat{\sigma}^1_j = \hat{\sigma}_j, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\hat{\sigma}\})$. This is because in the ex-post event $(v_1, \dots, v_N)$ such that $v_j = 3\delta, \forall j \neq i$ and $v_i = v'_i$, and $(a_1, \dots, a_N) \in \prod_{j \in N} A_j(\hat{\sigma}_j(v_j))$ such that $a_j = \delta, \forall j \in N$, player i obtains an expected payoff of $(v'_i - \delta)/N$ and hence, the bid of 2δ with a payoff of $v'_i - 2\delta$ is a better reply to a_{-i}. By repeating this argument we will reach a $\hat{\sigma}^l$ such that for all $j \in N$, $\hat{\sigma}^l_j(v_j, 0) = 1, \forall v_j \in [0, \delta]$, $\hat{\sigma}^l_j(v_j, \delta) = 1, \forall v_j \in [2\delta, 3\delta]$ and $\hat{\sigma}^l_j(v_j, 2\delta) = 1, \forall v_j \in [4\delta, 1]$. By construction, $\hat{\sigma}^l \in \tilde{R}^l_{FPA(RM)}(\{\hat{\sigma}\})$. Now, pick $v'_i \in [5\delta, 1]$. Let $\hat{\sigma}^{l+1}_i \in \Sigma_i(\hat{\sigma}^l_i)$ be such that $\hat{\sigma}^{l+1}_i(v'_i, 3\delta) > \hat{\sigma}^l_i(v'_i, 3\delta)$ and $\hat{\sigma}^{l+1}_i(v'_i, 2\delta) < \hat{\sigma}^l_i(v'_i, 2\delta)$. Then the strategy

35

profile $\hat{\sigma}^{l+1} = (\hat{\sigma}_1^{l+1}, \ldots, \hat{\sigma}_N^{l+1})$ such that $\hat{\sigma}_j^{l+1} = \hat{\sigma}_j^l, \forall j \neq i$ is in $\tilde{R}_{FPA(RM)}(\{\hat{\sigma}^l\})$ and hence, in $\tilde{R}_{FPA(RM)}^{l+1}(\{\sigma\})$. This is because in the ex-post event $(v_1, \ldots, v_N)$ such that $v_j = 4\delta, \forall j \neq i$ and $v_i = v_i'$, and $(a_1, \ldots, a_N) \in \prod_{j \in N} A_j(\hat{\sigma}_j^l(v_j))$ such that $a_j = 2\delta, \forall j \in N$, player i obtains an expected payoff of $(v_i' - 2\delta)/N$ and hence, the bid of 3δ with a payoff of $v_i' - 3\delta$ is a better reply to a_{-i}. By repeating this argument we will reach a $\hat{\sigma}^{l'}$ such that for all $j \in N$, $\hat{\sigma}_j^{l'}(v_j, 0) = 1, \forall v_j \in [0, \delta]$, $\hat{\sigma}_j^{l'}(v_j, \delta) = 1, \forall v_j \in [2\delta, 3\delta]$, $\hat{\sigma}_j^{l'}(4\delta, 2\delta) = 1$, and $\hat{\sigma}_j^{l'}(v_j, 3\delta) = 1, \forall v_j \in [5\delta, 1]$. By construction, $\sigma^{l'} \in \tilde{R}_{FPA(RM)}^{l'}(\{\hat{\sigma}\})$. It should be clear that the statement in the lemma can be proved after a finite number of repetitions of the above argument. $\square$

Proof of Proposition 4.11: We use the weighted cycle test to prove this proposition. However, we need to modify the weighted-adjacent-ex-post-better-reply graph of $FPA(RM)$ to incorporate the fact that in the event of a tie at the winning bid, the ex-post payoff of a bidder need not equal her expected payoff (where the expectation is with respect to the tie-breaking lottery). This is achieved by redefining $R_{FPA(RM)}$ as follows: $(\sigma_1', \ldots, \sigma_N') \in R_{FPA(RM)}(\sigma_1, \ldots, \sigma_N)$ if and only if

1. $(\sigma_1', \ldots, \sigma_N') \in \prod_{i \in N} \Sigma_i(\sigma_i)$ and

2. if $(\sigma_1', \ldots, \sigma_N') \neq (\sigma_1, \ldots, \sigma_N)$, then let $I = \{i | \sigma_i \neq \sigma_i'\}$. There must exist $(v_1, \ldots, v_N) \in V$ and $(a_1, \ldots, a_N) \in \prod_{i \in N} A_i(\sigma_i(v_i))$ such that for all $i \in I$, we have $v_i = v_i^{(\sigma_i, \sigma_i')}$, $a_i = a_i^{(\sigma_i, \sigma_i')}$ and the expected payoff of player i of valuation $v_i^{(\sigma_i, \sigma_i')}$ from bidding $a_i'^{(\sigma_i, \sigma_i')}$ against a_{-i} is greater than her payoff in some ex-post event (i.e., after the realization of the tie-breaking lottery) following her bid of $a_i^{(\sigma_i, \sigma_i')}$ against a_{-i}.

Now, use this redefined $R_{FPA(RM)}$ to modify the weighted-adjacent-ex-post-reply graph of $FPA(RM)$ by adding an edge from a strategy profile σ to $\sigma' \neq \sigma$ if and only if $\sigma' \in R_{FPA(RM)}(\sigma)$. The weight of this edge is the resistance $r(\sigma, \sigma')$ (note that we don't have to modify the definition of the resistance since it was already defined with respect to ex-post regrets). The weighted cycle test is also valid for this modified weighted-adjacent-ex-post-reply graph of $FPA(RM)$.

We prove by induction that a state σ is stochastically stable only if $\forall n = 0, \ldots, \frac{1}{\delta}$, the support of $\sigma_i(n\delta)$, $A_i(\sigma_i(n\delta)) \subseteq \left[\frac{(n-1)\delta}{2}, \frac{(n+1)\delta}{2}\right] \cap Z, \forall i \in N$.

First, we show that the statement is true for $n = 0$. Consider any state σ such that there exists a bidder of 0 valuation who plays a positive bid. Let z be the greatest positive bid played by any 0 valuation type of any bidder. Without loss of generality, let position i be such that $\sigma_i(0, z) > 0$. Let $\sigma_i^1 \in \Sigma_i(\sigma_i)$ be such that

$\sigma_i^1(0,0) > \sigma_i(0,0)$ and $\sigma_i^1(0,z) < \sigma_i(0,z)$. Then the strategy profile $\sigma^1 = (\sigma_1^1, \ldots, \sigma_N^1)$ such that $\sigma_j^1 = \sigma_j, \forall j \neq i$ is in $R_{FPA(RM)}(\{\sigma\})$. This is because in the ex-post event $(v_1, \ldots, v_N)$ such that $v_j = 0, \forall j \in N$ and $(a_1, \ldots, a_N) \in \prod_{j \in N} A_j(\sigma_j(0))$ such that $a_i = z$, bidder i obtains a negative payoff and hence, the bid of 0 is a better reply to a_{-i}. Hence, $r(\sigma, \sigma^1) < \frac{1}{\gamma}$ since only a single bidder of valuation 0 in population $\mathcal{P}_i$ switches from a positive bid to a bid of 0 and there exists an ex-post event in state σ in which this bidder has positive regret from bidding her positive bid instead of 0. Thus, in the weighted-adjacent-ex-post-better-reply graph, there exists an outgoing edge from σ to σ^1 with a weight of less than $\frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. Since $\sigma_i^1(0,0) > \sigma_i(0,0)$, it must be that there exist two states σ^l, σ^{l+1} such that $\sigma^{l+1} \in R_{FPA(RM)}(\sigma^l)$ and $\sigma_i^{l+1}(0,0) < \sigma_i^l(0,0)$. That is, there exists a bidder of valuation 0 in population $\mathcal{P}_i$ who switches from bidding 0 to some positive bid. However, there does not exist any a_{-i} such that bidder i of valuation 0 would strictly prefer to bid some positive bid instead of 0. Hence, $r(\sigma^l, \sigma^{l+1}) \geq \frac{1}{\gamma}$. Thus, the state σ passes the wighted cycle test and hence, it is not stochastically stable.

Now, suppose the statement is true for all $n' \leq n$. We argue that it is also true for $n+1$. Consider a state σ such that $\forall j \in N$ and $v_j \leq n\delta$, we have $A_j(\sigma_j(v_j)) \subseteq \left[\frac{v_j-\delta}{2}, \frac{v_j+\delta}{2}\right] \cap Z$ but there exists a bidder of valuation $(n+1)\delta$ in population $\mathcal{P}_i$ who bids $z \notin \left[\frac{n\delta}{2}, \frac{(n+2)\delta}{2}\right] \cap Z$.

First suppose $z > \frac{(n+2)\delta}{2}$. Let $\sigma_i^1 \in \Sigma_i(\sigma_i)$ be such that $\sigma_i^1((n+1)\delta, \delta) > \sigma_i((n+1)\delta, \delta)$ and $\sigma_i^1((n+1)\delta, z) < \sigma_i((n+1)\delta, z)$. Consider the ex-post event $(v_1, \ldots, v_N)$ such that $v_j = 0, \forall j \neq i$ and $v_i = (n+1)\delta$, and $(a_1, \ldots, a_N) \in \prod_{j \in N} A_j(\sigma_j(v_j))$ such that $a_j = 0, \forall j \neq i$ and $a_i = z$. Then bidding δ is a better reply to a_{-i} than z and bidder i's ex-post regret is $z - \delta$. Hence, $r(\sigma, \sigma^1) \leq \frac{1}{z-\delta} < \frac{1}{\gamma}$ since only a single bidder of valuation $(n+1)\delta$ in population $\mathcal{P}_i$ switches from the bid of z to the bid of δ and there exists an ex-post event in state σ in which this bidder has positive ex-post regret of $z - \delta$ from bidding z instead of δ. Thus, in the weighted-adjacent-ex-post-better-reply graph of $FPA(RM)$, there exists an outgoing edge from σ to σ^1 with a weight of less than $\frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. Since $\sum_{z' \geq z} \sigma_i^1((n+1)\delta, z') < \sum_{z' \geq z} \sigma_i((n+1)\delta, z')$, it must be that there exist two states σ^l, σ^{l+1} such that $\sigma^{l+1} \in R_{FPA(RM)}(\sigma^l)$ and $\sum_{z' \geq z} \sigma_i^{l+1}((n+1)\delta, z') > \sum_{z' \geq z} \sigma_i^l((n+1)\delta, z')$. That is, there exists a bidder of valuation $(n+1)\delta$ in population $\mathcal{P}_i$ who switches from bidding $\tilde{z} < z$ to some $\hat{z} \geq z$. However, the bidder i of valuation $(n+1)\delta$ can gain a positive amount from increasing her bid from $\tilde{z}$ to $\hat{z}$ against any a_{-i} only if $\hat{z} < (n+1)\delta$ and the maximum such gain

equals $(n+1)\delta-\hat{z}$ (when she turns from loosing at $\tilde{z}$ to winning at $\hat{z}$). Since $\hat{z} > \frac{(n+2)\delta}{2}$, the maximum gain from such a switch is less than $\frac{n\delta}{2}$. Hence, $r(\sigma^l, \sigma^{l+1}) > \frac{2}{n\delta} > \frac{1}{z-\delta}$ since $z > \frac{(n+2)\delta}{2}$. Thus, the state σ passes the wighted cycle test and hence, it is not stochastically stable.

Finally, suppose $z < \frac{n\delta}{2}$. This is possible only if $n \geq 1$. Let z^* be the least bid greater than or equal to $\frac{n\delta}{2}$. Let $\sigma_i^1 \in \Sigma_i(\sigma_i)$ be such that $\sigma_i^1((n+1)\delta, z^*) > \sigma_i((n+1)\delta, z^*)$ and $\sigma_i^1((n+1)\delta, z) < \sigma_i((n+1)\delta, z)$.

Case 1: n is odd. Then $z \leq \frac{(n-1)\delta}{2}$ and $z^* = \frac{(n+1)\delta}{2}$. Pick a $j \neq i$ and $v_j = (n-1)\delta$. By the induction hypothesis, there exists a $z' \in \left[\frac{(n-2)\delta}{2}, \frac{n\delta}{2}\right] \cap Z$ such that $\sigma_j(v_j, z') > 0$. Then $z' = \frac{(n-1)\delta}{2}$. Consider the ex-post event $(v_1, \ldots, v_N)$ such that $v_k = 0, \forall k \neq i, j$, $v_j = (n-1)\delta$ and $v_i = (n+1)\delta$, and $(a_1, \ldots, a_N) \in \prod_{k \in N} A_k(\sigma_k(v_k))$ such that $a_k = 0, \forall k \neq i, j, a_j = z'$ and $a_i = z$. Since $z \leq z'$, with a positive probability, bidder i looses the auction. Therefore, in this ex-post stage, bidding z^* is a better reply to a_{-i} than z and bidder i's ex-post regret is $(n+1)\delta - z^*$. Hence, $r(\sigma, \sigma^1) \leq \frac{1}{(n+1)\delta - z^*} < \frac{1}{\gamma}$ since only a single bidder of valuation $(n+1)\delta$ in population $\mathcal{P}_i$ switches from the bid of z to the bid of z^* and there exists an ex-post event in state σ in which this bidder has positive ex-post regret of $(n+1)\delta - z^*$ from bidding z instead of z^*. Thus, in the weighted-adjacent-ex-post-better-reply graph of $FPA(RM)$, there exists an outgoing edge from σ to σ^1 with a weight of less than $\frac{1}{\gamma}$. Now, consider any weighted cycle containing this outgoing edge. Since $\sigma_i^1((n+1)\delta, z^*) > \sigma_i((n+1)\delta, z^*)$, it must be that there exist two states σ^l, σ^{l+1} such that $\sigma^{l+1} \in R_{FPA(RM)}(\sigma^l)$ and $\sigma_i^{l+1}((n+1)\delta, z^*) < \sigma_i^l((n+1)\delta, z^*)$. That is, there exists a bidder of valuation $(n+1)\delta$ in population $\mathcal{P}_i$ who switches from bidding z^* to some $\hat{z} \neq z^*$. Suppose $\hat{z} > z^*$. The bidder i of valuation $(n+1)\delta$ can gain a positive amount from increasing her bid from z^* to $\hat{z}$ against any a_{-i} only if $\hat{z} < (n+1)\delta$ and the maximum such gain equals $(n+1)\delta - \hat{z} < (n+1)\delta - z^*$ (when she turns from loosing at z^* to winning at $\hat{z}$). Next, suppose $\hat{z} < z^*$. The bidder i of valuation $(n+1)\delta$ can gain a positive amount from decreasing her bid from z^* to $\hat{z}$ against any a_{-i} only if $\hat{z} > \max_{j \neq i} a_j$ and the maximum such gain equals $z^* - \delta < (n+1)\delta - z^*$ (when turns from winning at z^* to winning at the bid of δ). Hence, $r(\sigma^l, \sigma^{l+1}) > \frac{1}{(n+1)\delta - z^*} \geq r(\sigma, \sigma^1)$. Thus, the state σ passes the wighted cycle test and so, it is not stochastically stable.

Case 2: n is even. Then $n \geq 2$, $z \leq \frac{(n-2)\delta}{2}$ and $z^* = \frac{n\delta}{2}$. Pick a $j \neq i$ and $v_j = (n-2)\delta$. By the induction hypothesis, there exists a $z' \in \left[\frac{(n-3)\delta}{2}, \frac{(n-1)\delta}{2}\right] \cap Z$ such that $\sigma_j(v_j, z') > 0$. Then $z' = \frac{(n-2)\delta}{2}$. Consider the ex-post event $(v_1, \ldots, v_N)$ such that $v_k = 0, \forall k \neq i, j$, $v_j = (n-2)\delta$ and $v_i = (n+1)\delta$, and $(a_1, \ldots, a_N) \in \prod_{k \in N} A_k(\sigma_k(v_k))$ such that $a_k = 0, \forall k \neq i, j, a_j = z'$ and $a_i = z$. Since $z \leq z'$, with

38

a positive probability, bidder i looses the auction. Therefore, in this ex-post stage, bidding z^* is a better reply to a_{-i} than z and bidder i's ex-post regret is $(n+1)\delta - z^*$. Hence, $r(\sigma, \sigma^1) \leq \frac{1}{(n+1)\delta - z^*} < \frac{1}{\gamma}$. Thus, in the weighted-adjacent-ex-post-better-reply graph of $FPA(RM)$, there exists an outgoing edge from σ to σ^1 with a weight of less than $\frac{1}{\gamma}$. Like in case 1, we can easily argue that in any weighted cycle containing this outgoing edge, there exists an edge with weight greater than $r(\sigma, \sigma^1)$. Hence, σ is not stochastically stable. □

References

Dekel, E., Fudenberg, D., Levine, D. K., 2004. Learning to Play Bayesian Games. Games and Economic Behavior 46, 282-303.

Ely, J. C., Sandholm, W. H., 2005. Evolution in Bayesian Games I: Theory. Games and Economic Behavior 53, 83-109.

Fudenberg, D., Levine, D. K., 1998. The Theory of Learning in Games. Cambridge: MIT Press.

Hart, S., 2005. Adaptive Heuristics. Econometrica 73, 1401-1430.

Hart, S., Mas-Colell, A., 2003. Uncoupled Dynamics Do Not Lead to Nash Equilibrium. American Economic Review 93, 1830-1836.

Hon-Snir, S., Monderer, D., Sela, A., 1998. A Learning Approach to Auctions. Journal of Economic Theory 82, 65-88.

Hyafil, N., Boutilier, C., 2004. Regret Minimizing Equilibria and Mechanisms for Games with Strict Type Uncertainty. In: Chickering, M., Halpern, J. (Eds.). Uncertainty in Artificial Intelligence: Proceedings of the 20th Conference. Arlington: AUAI Press, 268-277.

Jensen, M., Sloth, B., Whitta-Jacobsen, H. J., 2005. The Evolution of Conventions Under Incomplete Information. Economic Theory 25, 171-185.

Kandori, M., Mailath, G., Rob, R., 1993. Learning, Mutation, and Long-Run Equilibria in Games. Econometrica 61, 29-56.

Linhart, P. B., Radner, R. 1989. Minimax-Regret Strategies for Bargaining over Several Variables. Journal of Economic Theory 48, 152-178.

Ritzberger, K., Weibull, J. W., 1995. Evolutionary Selection in Normal-Form Games. Econometrica 63, 1371-1399.

Sandholm, W. H., 2009. Evolutionary Game Theory. In: Meyers, R. (Ed.). Encyclopedia of Complexity and Systems Science. New York: Springer, 3176-3205.

Saran, R., Serrano, R., 2007. The Evolution of Bidding Behavior in Private-Values Auctions and Double Auctions. Working Paper 2007-01, Department of Economics, Brown University.

Saran, R., Serrano, R., 2010. Regret Matching with Finite Memory. Mimeo, Department of Economics, Brown University.

Savage, L. J., 1951. The Theory of Statistical Decision. Journal of American Statistical Association 46, 55-67.

Spulber, D. F., 1995. Bertrand Competition when Rivals'Costs are Unknown. Journal of Industrial Economics 43, 1-11.

Young, H. P., 1993. The Evolution of Conventions. Econometrica 61, 57-84.

Young, H. P., 1998. Individual Strategy and Social Structure. Princeton: Princeton University Press.

Young, H. P., 2004. Strategic Learning and its Limits. Oxford: Oxford University Press.