On the Impossibility of Achieving No Regrets in Repeated Games*

Karl Schlag
University of Vienna†

Andriy Zapechelnyuk
Queen Mary, University of London‡

September 15, 2011

Abstract

Regret-minimizing strategies for repeated games have been receiving increasing attention in the literature. These are simple adaptive behavior rules that lead to no regrets and, if followed by all players, exhibit nice convergence properties: the average play converges to correlated equilibrium, or even to Nash equilibrium in certain classes of games. However, the no-regret property relies on a strong assumption that each player treats her opponents as unresponsive and fully ignores the opponents' possible reactions to her actions. We show that if at least one player is slightly responsive, it is impossible to achieve no regrets, and convergence results for regret minimization with responsive opponents are unknown.

Keywords: Adaptive strategies, regret minimization, regret matching

JEL classification numbers: C73, D81, D83

---

*We thank Jörg Oechssler and an anonymous referee for valuable comments that led to a significant improvement of the paper.

†Department of Economics, University of Vienna, Hohenstaufengasse 9, 1010 Vienna, Austria. *E-mail:* karl.schlag $\alpha\tau$ univie.ac.at

‡*Corresponding author.* School of Economics and Finance, Queen Mary, University of London, Mile End Road, London E1 4NS, UK. *E-mail:* a.zapechelnyuk $\alpha\tau$ qmul.ac.uk

1. Introduction. In a repeated interaction, an individual follows a *regret-minimizing* strategy if, loosely speaking, she reinforces those actions that she regrets not having played enough in the past. A particularly simple strategy is *regret matching*, which is defined by the following rule:

> Switch next period to a different action with a probability that is *proportional* to the *regret* for that action, where regret is defined as the increase in payoff had such a change always been made in the past (Hart and Mas-Colell, 2000; Hart, 2005).

This strategy, in particular, has the property "never change a winning team," in other words, do not switch to a different action, as long as the current action keeps being a best reply to the observed (average) actions of the opponents.

Regret-minimizing strategies that lead to "no regrets" irrespective of what the opponents play, called *no-regret strategies*, received a lot of attention in the recent literature.[1] The main value of these strategies is that they are simple adaptive behavior rules that are neither computationally demanding nor relying on common knowledge assumptions and yet exhibiting nice convergence properties. If all players follow no-regret strategies, their average joint play converges to the set of correlated equilibria or to the Hannan set[2], depending on the notion of regret in use (Hart and Mas-Colell 2000; see also Lehrer 2003, Cesa-Bianchi and Lugosi 2006); or even to Nash equilibria on certain classes of games (Hart and Mas-Colell 2003; Marden, Arslan, and Shamma 2007).

[1]A non-exhaustive list includes Littlestone and Warmuth (1994), Fudenberg and Levine (1995), Foster and Vohra (1998), Foster and Vohra (1999), Freund and Schapire (1999), Hart and Mas-Colell (2000), Hart and Mas-Colell (2001), Hart and Mas-Colell (2003), Lehrer (2003), Young (2004), Cesa-Bianchi and Lugosi (2003), Cesa-Bianchi and Lugosi (2006), Lehrer and Solan (2009).

[2]The Hannan set of a game is the set of all mixed action profiles that satisfy Hannan's (1957) no-regret condition. It is also known as the set of *coarse correlated equilibria* first appeared in Moulin and Vial (1978), but explicitly defined as a solution concept by Young (2004, Ch.3).

In this note we raise the question of validity of the regret minimization objective in the context of games. On the one hand, according to the notions of regret used in the literature, an individual who contemplates whether she could have done better by having played a particular action more often in the past does not take into account the effect of her actions on the subsequent behavior of her opponent. This is perfectly fine in a decision making environment, but *not* in a game, where, by definition, players are responsive to their opponents' behavior. We show by example that failure to take the opponent's responsiveness into account may lead to unrealistic behavior.[3]

On the other hand, if we extend the notion of regret to take into account the above mentioned effect, then it becomes impossible to guarantee no regrets, even against a severely restricted set of the opponent's strategies. We show that if an opponent is slightly responsive to the player's past behavior, the maximum regret need not converge to zero. Consequently, even if all players play regret-minimizing strategies (such as Hart and Mas-Colell's (2000) regret matching) with respect to this extended notion of regret, their regrets need not vanish in the long run, and consequently, the known convergence results are not guaranteed.

2. Regrets. Consider a finite two-player game, with players named Alice and Bob.[4] Let A and B be sets of actions of Alice and Bob, respectively, and let $u : A \times B \to \mathbb{R}$ be Alice's payoff function. The game is played repeatedly in time periods $t = 1, 2, \ldots$, in which players choose actions (a_t, b_t). The history of realized actions is observable for both players.

[3]This problem is recognized in the computer science literature. Farias and Megiddo (2004) and Cesa-Bianchi and Lugosi (2006, Ch.7.11) show that regret minimizing strategies fail to lead to the cooperative outcome in a repeated prisoner's dilemma. Our example is different and, as we believe, has a value on its own, as it illuminates failure to learn the Pareto dominant equilibrium of a *one-shot game*, whereas the above literature shows failure to learn playing strictly dominated actions.

[4]Bob can be considered as a set of players, so the arguments presented below trivially extend to n-player games.

Denote by $\bar{U}_T(a,b)$ the average payoff of Alice up to period T,

$$\bar{U}_T(a,b) = \frac{1}{T}\sum_{t=1}^{T} u(a_t, b_t),$$

and denote by $U_T(a_{(a^*|a')}, b)$ the average payoff that Alice would have obtained had she played a' instead of the reference action a^* every time in the past when she actually played a^*,

$$U_T(a_{(a^*|a')}, b) = \frac{1}{T}\sum_{t=1}^{T} w_t(a'),$$

where

$$w_t(a') = \begin{cases} u(a', b_t), & \text{if } a_t = a^*, \\ u(a_t, b_t), & \text{if } a_t \neq a^*. \end{cases}$$

Alice's *regret* $r_T(a', a^*; a, b)$ for choosing action a^* instead of action a' after T periods is defined as the excess of $U_T(a_{(a^*|a')}, b)$ over $\bar{U}_T(a,b)$,

$$r_T(a', a^*; a, b) = U_T(a_{(a^*|a')}, b) - \bar{U}_T(a,b).$$

The objective of the previous literature has been to identify strategies for Alice that guarantee no regrets in the long run. More specifically, let $h_T = ((a_1, b_1), \ldots, (a_T, b_T))$ denote the history of play up to T, and let $\mathcal{H}$ be the set of all finite histories. Alice has a no-regret strategy $\alpha : \mathcal{H} \to \Delta(A)$ if $\limsup_{T\to\infty} r_T(a', a^*; a, b) \leq 0$ holds almost surely under α for all deterministic sequences b and all pairs of actions (a^*, a').

According to the above definition of regret[5], Alice evaluates her regret for choosing action a^* instead of action a' by contemplating how much higher

[5]Specifically, we have been considering *conditional regrets*. The *unconditional regret* of Alice for an action a' refers to the difference in her average payoff had she always chosen a' instead of her actual past play. "No conditional regret" implies "no unconditional regret", but not vice versa, unless Alice has only two actions.

payoff, on average, she could have received had she played a' in every past period when she actually played a^*, assuming that *the play of the opponents would have remained unchanged*. This definition is plausible in the context of decision making, when an individual's actions have no effect on the opponent, who can be perceived as an abstract environment. It is much less appealing if the individual is engaged in a game, where the opponent's future play can be responsive to the individual's present actions.

	Bob	
Alice	L	R
L	$1, 1$	$0, 0$
R	$0, 0$	$100, 100$

Figure 1

3. An example. For illustration, consider the following coordination game (Fig. 1). Suppose that the observed play up to period $T \geq 2$ is $((a_1, b_1), (a_2, b_2), \ldots, (a_T, b_T)) = ((L, L), (L, L), \ldots, (L, L))$. Given this history, from the perspective of Alice, playing L is a best reply to the average *realized* play of Bob.

Does Alice have regret for action R? Not according to the above definition. Looking at the observed sequence of play, Alice takes the actions of Bob as given and realizes that had she chosen R every time she chose L, she would have gotten a lower payoff.

This argument relies on the assumption that the sequence of Bob's actions does not depend on what Alice plays. In other words, Bob's action in a given period does not depend on the previous choices of Alice. However, if Bob's strategy is *adaptive*, in the sense that Bob's choices depend on Alice's previous actions, then it is not clear whether or not Alice could not have done better. For instance, if Bob's strategy is to choose a best response to what Alice did in the last round, then Alice would regret not choosing R.

To summarize: *Could Alice have done better by having switched to R?*

(I) No, if Bob's strategy is *independent* of Alice's actions.

(II) Possibly, if Bob's strategy is *adaptive*.

As argued above, the definition of regret implicitly relies on the assumption that Bob's strategy is independent of Alice's actions. However, in game situations players can react to what the opponents do. Hence, the evaluation of the regret of not choosing R depends on how the opponent would have reacted to this change. This reaction has to be included into the definition of regret.

Note that the above discussion does not require that Alice knows Bob's payoffs. It is only Bob's behavior that Alice has to be concerned with, in particular, whether or not Bob can condition his choices on what Alice does and react as described above.

4. Regrets against history dependent behavior. Above we consider regrets of Alice when facing a given sequence of actions chosen by Bob. In other words, the strategy of Bob was independent of Alice's actions. Let us now adjust our definition to allow for Bob to be adaptive, namely, to react to what Alice has chosen in the past. To increase generality we will also allow Bob to choose mixed actions, and hence to use a mixed strategy.

Denote by $h_T = ((a_1, b_1), \ldots, (a_T, b_T))$ the history of play up to T, and let $\mathcal{H}$ be the set of all finite histories. Let $\alpha : \mathcal{H} \to \Delta(A)$ and $\beta : \mathcal{H} \to \Delta(B)$ be strategies of Alice and Bob, respectively, that prescribe mixed actions for every history $h_t \in \mathcal{H}$. Denote by $U_T(\alpha, \beta)$ the expected average payoff of Alice up to period T when she plays α against Bob playing β,

$$U_T(\alpha, \beta) = E_{(\alpha,\beta)} \left[\frac{1}{T} \sum_{t=1}^{T} u(a_t, b_t) \right],$$

where the expectation is taken with respect to the probability measure over

$\mathcal{H}$ induced by (α, β).

Fix Alice's strategy α. Denote by $\alpha_{(a^*|a')}$ the strategy obtained from α by replacement of a^* by a' in all periods where the realized action of α is a^*. Formally, for every history $h \in \mathcal{H}$ let

$$\alpha_{(a^*|a')}(h)[a^*] = 0, \quad \text{and}$$
$$\alpha_{(a^*|a')}(h)[a'] = \alpha(h)[a^*] + \alpha(h)[a'],$$

where $\alpha(h)[k]$ denotes the probability that $\alpha(h)$ assigns to action $k \in A$.

For a given strategy β of Bob, $U_T(\alpha_{(a^*|a')}, \beta)$ is the expected average payoff that Alice would have obtained had she played a' every time in the past when her strategy α stipulated to play a^*, and when at every stage $t \leq T$ Bob would have responded according to β to the new history.

Let $\mathcal{B}$ be a set of Bob's feasible strategies, and let $\beta \in \mathcal{B}$. Then Alice's regret for choosing a^* instead of a' is given by

$$\rho_T(a', a^*; \alpha, \beta) = U_T(\alpha_{(a^*|a')}, \beta) - \bar{U}_T(\alpha, \beta).$$

If $\rho_T(a', a^*; \alpha, \beta) \leq 0$ for all $\beta \in \mathcal{B}$, then Alice can conclude that she could not have done better by switching a^* to a' in the past, no matter what is the actual strategy of Bob.

A strategy of Alice is called a *no-regret strategy against* $\mathcal{B}$ if it guarantees that Alice's regrets become non-positive in the limit for every strategy of Bob in $\mathcal{B}$,

$$\limsup_{T\to\infty} \rho_T(a', a^*; \alpha, \beta) \leq 0 \quad \text{for all } \beta \in \mathcal{B} \text{ and all } a', a^* \in A.$$

We hasten to point out that, apart from the use of expectations, the definition of no-regret strategy for adaptive opponents is identical to that for non-adaptive opponents. In particular, we do not introduce a new notion of regret, we only adapt the notion to a richer set of strategies of Bob.

It is known that there exist no-regret strategies against an unresponsive opponent as considered in Section 2, i.e., when $\mathcal{B}$ contains only deterministic sequences (or distributions over such sequences) (e.g., Hannan, 1957; Hart and Mas-Colell, 2000, 2001; Cesa-Bianchi and Lugosi, 2003). Yet, as we show below, a minimum of adaptiveness of Bob's strategies to Alice's past actions leads to an impossibility result.

Bob's strategy is called *q-fictitious play* if in every period $t = 2, 3, \ldots$, with probability $1-q$ Bob repeats his last-period action, and with probability q he best-replies to Alice's average past play. The initial play of Bob is arbitrary.

For some $\varepsilon > 0$ denote by $\mathcal{B}_\varepsilon$ the set of q-fictitious play strategies with $q \in [0, \varepsilon]$. In particular, $\mathcal{B}_\varepsilon$ contains non-adaptive strategies where Bob plays a constant action (0-fictitious play).

The next proposition shows that no strategy can guarantee the maximum regret to converge to zero if Alice cannot exclude the possibility that her opponent is responsive, even if the degree of responsiveness is arbitrarily small.

Proposition. *There exists a game such that, for every $\varepsilon > 0$, there does not exist a no-regret strategy against $\mathcal{B}_\varepsilon$.*

Before proving the proposition, let us briefly explain the intuition behind it. Assume that Bob plays L in period one. In order to guarantee no regrets, Alice needs to identify whether Bob is non-adaptive ($q = 0$) and thus playing constant action L, or he is adaptive ($q > 0$) and thus able to coordinate on the Pareto superior equilibrium (R, R). In the former case, Alice should always respond by action L, whereas in the latter case Alice can guarantee convergence to equilibrium (R, R) with probability one by always playing R. To see this, observe that since $q > 0$, the probability that Bob has never played fictitious play by period t is $(1-q)^t$, which converges to zero as $t \to \infty$. Hence, with probability one Bob will eventually best-reply to Alice's past average play. If Alice has been playing R frequently enough, Bob will switch to R as well, and then the joint play locks in (R, R) forever. However, the problem is that *Alice can never be confident enough that Bob is non-adaptive, no matter how long she observes Bob playing L.* Since $q > 0$ can be arbitrarily small, the fact that Bob has always played L conveys no information about Bob's adaptiveness. That is to say, from Alice's perspective, Bob's types $q = 0$ and $q > 0$ are statistically indistinguishable.

Proof. Consider the coordination game described earlier (Fig. 1). Fix $\varepsilon > 0$ and suppose that Bob plays q-fictitious play, $\beta^q \in \mathcal{B}_\varepsilon$, $q \in [0, \varepsilon]$, and let his initial action be L. More specifically, in every period $t \geq 2$, with probability q Bob chooses action R if Alice has played R at least $1/100$ fraction of time so far, and otherwise Bob chooses L; with probability $1 - q$ Bob repeats his last-period action.

Observe that if Alice knew that $q = 0$, then her best reply would be to always play L, since Bob is non-adaptive and repeats L forever, so $\bar{U}_T = u(L, L) = 1$. On the other hand, if she knew that $q > 0$, then her best reply would be to always play R, since eventually, with probability 1, Bob would switch to R after observing Alice's past average play being R, and the further play would be locked on (R, R) forever, so $\bar{U}_T \to u(R, R) = 100$ as $T \to \infty$. Note that in order to derive Alice's regret for choosing action, say, R instead of L, one needs to replace Alice's action L by R in every instance where she plays L. Since there are only two actions, it means to compare the performance of strategy α with constant play of R. Thus the two constant strategies (always R and always L) are our benchmarks relative to which Alice will measure her regret. The task of Alice is to design a strategy that, without knowing $q = 0$ or $q > 0$, will result in limit average payoffs close to 1 in the former case and to 100 in the latter case. We will show that this construction is impossible.

Suppose by contradiction that there exists a no-regret strategy for Alice against $\mathcal{B}_\varepsilon$. Note that after the first time, t, where Bob played R, Alice has a no-regret strategy for the subgame on $t + 1, t + 2, \ldots$, by playing R constantly from $t + 1$ on. To see this, observe that in every period after t, Bob either repeats his last action, R, or best-replies to Alice's average play. The best reply action is also R, since it has been his best-reply in period t and Alice has played only R since then. Alice's payoff will be constantly 100 from period $t+1$ on, and hence, she will have no regrets. Thus, a no-regret strategy (if it exists) can be fully described by Alice's play in every period t, so long as Bob plays L; and it stipulates to play R constantly after the first time Bob played R.

Let z_t^* be the frequency with which Alice chose action R in $\{a_1, \ldots, a_t\}$. Consider the subsequence of periods, $\{t_s\}$, such that $z_{t_s}^* \geq 1/100$ and Bob has chosen L constantly up to t_s. These are the periods where Bob would have played R had he taken the best-reply action.

First, suppose that this subsequence is finite, i.e., there is a number S such that $|\{t_s\}| = S$. Let us evaluate the expected payoff for Alice when Bob follows q ficitious play with $q > 0$. The probability that Bob never plays R up to t_S is equal to $(1-q)^S$. In this event the play will be locked on (L, L) forever, thus Alice's average payoff $\bar{U}_T$ will approach $u(L, L) = 1$. With the complementary probability $1 - (1-q)^S$, Bob plays R before or on t_S. From that period on the play is locked on (R, R), thus Alice's average payoff $\bar{U}_T$ will approach $u(R, R) = 100$. The expected payoff (from the perspective of period zero) will therefore approach

$$\lim_{T\to\infty} \bar{U}_T = (1 - (1-q)^S) \cdot 100 + (1-q)^S \cdot 1 = 100 - 99(1-q)^S.$$

Thus, Alice's regret for not playing R constantly is

$$\lim_{T\to\infty} (U_T(R) - \bar{U}_T) = 99(1-q)^S.$$

It is straightforward to see that, for a given S, this regret is bounded away from zero for every sufficiently small q. For example, for every $q \leq 9/(S+9)$ the regret is at least $1/100$.[6]

[6] Inequality $99(1-q)^S \geq 1/100$ is equivalent to $S\ln(1-q) \geq -\ln 9900$. Since $\ln(1-q) \geq$

Alternatively, suppose that subsequence $\{t_s\}$ is infinite. Let us evaluate the expected payoff of Alice when Bob is non-adaptive ($q = 0$), so that he never plays R. In every period t_s, the frequency of action R in Alice's past play is at least $1/100$. Thus,

$$\bar{U}_{t_s} \leq \frac{99}{100}u(L,L) + \frac{1}{100}u(R,L) = \frac{99}{100}\cdot 1 + \frac{1}{100}\cdot 0 = \frac{99}{100}$$

Hence Alice's regret for not playing L constantly is

$$\limsup_{T\to\infty}(U_T(L) - \bar{U}_T) \geq \limsup_{s\to\infty}(U_{t_s}(L) - \bar{U}_{t_s}) \geq 1 - 99/100 = 1/100.$$

It follows that no matter what Alice plays, there exists a strategy in $\mathcal{B}_\varepsilon$ of Bob such that lim sup of Alice's regret for one of the constant actions is bounded away from zero. $\square$

One may wish to evaluate regret without taking expectations, in the spirit of an ex-post perspective, when looking back at what has happened. It is as if one only evaluates regret against Bob who is choosing some pure strategy. Our proof above does not show that such a no "ex-post" regret strategy fails to exist, as it relied on deriving regret when Bob chooses a mixed strategy. However, it can easily be adapted. We say that Bob plays a *deterministic q-fictitious play* for some $q \in [0,1]$ if there is a deterministic subsequence of periods where Bob best-replies to Alice's past play, with the property that up to every period T the fraction of periods where Bob has best-replied *does not exceed* q. It is then easy to see that if $\mathcal{B}_\varepsilon$ is the set of deterministic q-fictitious play strategies with $q \leq \varepsilon$, then a no-regret strategy for Alice does not exist for any $\varepsilon > 0$.

Alternatively, the following example shows that there does not exist a no-regret strategy for Alice against Bob using one of three simple pure strategies. The key to this example will be that Bob uses *trigger strategies.* Consider

$-q/(1-q)$ and $\ln 9900 \geq 9$, the above inequality holds if $Sq/(1-q) \leq 9$. Rearranging the terms yields $q \leq 9/(S+9)$.

	Bob		
Alice	L	M	R
L	$2, 2$	$1, 1$	$0, 0$
R	$0, 0$	$1, 1$	$2, 2$

Figure 2

the game on Fig. 2 and suppose that the set of strategies of Bob includes the following:

(***non-adaptive***) Bob constantly plays M.

(***adaptive-**L*) Bob starts with M. Then, if Alice played L in the initial period, then Bob will play L from period 2 forever, otherwise he will play M forever.

(***adaptive-**R*) the same as *adaptive-*L except L is replaced by R.

In this game, Alice's long-run average payoff is determined entirely by Bob's type and Alice's initial action, since Bob's actions are constant from period 2 on. Now observe that no matter what Alice plays in period 1, there is a type of Bob, either *adaptive-L* or *adaptive-R*, that would make her regret for action L or R, respectively, in all subsequent periods. Indeed, if Alice chooses, for instance, L in the first period and Bob's type is *adaptive-R*, then the following play of Bob will be constantly M, and Alice's average payoff will be 1. However, Alice could have obtained the average payoff of 2, had she started her play with R.

5. Conclusion. To sum up, the notion of regret used in the literature is not satisfactory in the context of repeated games as it fails to take into account possible reaction of opponents to changes in one's actions. We define an extended notion of regret, with respect to opponents' strategies (rather than realized actions) and show that in this case no-regret strategies need not exist when the opponents are adaptive. Two examples provide the intuition for this result: the regrets persist because the opponent's strategy cannot be statistically identified (as in the former example) or because the opponent uses *trigger* strategies, where an early decision of the player (which is payoff-relevant for the the entire infinitely repeated interaction) has to be made when the player has not been yet informed about the opponent's strategy. The first example, in fact, shows that the no-regret property of regret minimizing strategies is not robust, as it fails to hold even when there is only a small probability that the opponent is adaptive. All arguments extend to n-player games, where it suffices that one player is adaptive.

We conclude that the existing no-regret strategies should be used with caution in the context of repeated games. They are appropriate if players are boundedly rational and assume that their opponents are non-adaptive. However, more realistically, if players understand that their own behavior may influence others', then the no-regret property cannot be achieved.

References

Cesa-Bianchi, N. and G. Lugosi (2003). Potential-based algorithms in on-line prediction and game theory. *Machine Learning 51*, 239–261.

Cesa-Bianchi, N. and G. Lugosi (2006). *Prediction, Learning, and Games.* Cambridge University Press.

Farias, D. P. and N. Megiddo (2004). How to combine expert (or novice) advice when actions impact the environment. In *Advances in Neural Information Processing Systems*, Volume 16. MIT Press.

Foster, D. and R. Vohra (1998). Asymptotic calibration. *Biometrika 85*, 379–390.

Foster, D. and R. Vohra (1999). Regret in the online decision problem. *Games and Economic Behavior 29*, 7–35.

Freund, Y. and R. Schapire (1999). Adaptive game playing using multiplicative weights. *Games and Economic Behavior 29*, 79–103.

Fudenberg, D. and D. Levine (1995). Consistency and cautious fictitious play. *Journal of Economic Dynamics and Control 19*, 1065–1089.

Hannan, J. (1957). Approximation to Bayes risk in repeated play. In M. Dresher, A. W. Tucker, and P. Wolfe (Eds.), *Contributions to the Theory of Games, Vol. III*, Annals of Mathematics Studies 39, pp. 97–139. Princeton University Press.

Hart, S. (2005). Adaptive heuristics. *Econometrica 73*, 1401–1430.

Hart, S. and A. Mas-Colell (2000). A simple adaptive procedure leading to correlated equilibrium. *Econometrica 68*, 1127–1150.

Hart, S. and A. Mas-Colell (2001). A general class of adaptive procedures. *Journal of Economic Theory 98*, 26–54.

Hart, S. and A. Mas-Colell (2003). Continuous-time regret-based dynamics. *Games and Economic Behavior 45*, 375–394.

Lehrer, E. (2003). A wide range no-regret theorem. *Games and Economic Behavior 42*, 101–115.

Lehrer, E. and E. Solan (2009). Approachability with bounded memory. *Games and Economic Behavior 66*, 995–1004.

Littlestone, N. and M. Warmuth (1994). The weighted majority algorithm. *Information and Computation 108*, 212–261.

Marden, J. R., G. Arslan, and J. S. Shamma (2007). Regret based dynamics: convergence in weakly acyclic games. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS07)*, Honolulu, Hawaii, USA, pp. 194–201.

Moulin, H. and J. P. Vial (1978). Strategically zero-sum games: the class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory 7*, 201–221.

Young, H. P. (2004). *Strategic Learning and Its Limits.* Oxford University Press.