Limits to Rational Learning

Yehuda John Levy[*][†]

August 13, 2015

Abstract

A long-standing open question raised in the seminal paper of Kalai and Lehrer (1993) is whether or not the play of a repeated game, in the rational learning model introduced there, must eventually resemble the play of *exact* equilibria, and not just the play of *approximate* equilibria as demonstrated there. This paper shows that play may remain distant – in fact, mutually singular – from the play of any equilibrium of the repeated game. We further show that the same inaccessibility holds in Bayesian games, where the play of a Bayesian equilibrium may continue to remain distant from the play of any equilibrium of the true game.

JEL Classification: C73, D83, C65

1 Introduction

The premise of *rational learning* is that decision-making agents update their beliefs about what other agents will do based on the actions that they have observed. The seminal work of [Blackwell and Dubins (1962)] shows that when a single agent learns rationally in this way, even if his prior beliefs about the process are incorrect but do contain a minimal amount of truth, his posterior beliefs will eventually lead to true beliefs about the process. The staple work in game theory which incorporates this paradigm into the multi-agent setting is [Kalai and Lehrer (1993)], in which agents both learn and also try to maximise their utility in a repeated game. Their work studies the question of whether the beliefs of the agents not only merge, but converge to beliefs induced by the *Nash equilibria* of the repeated game. As it turns out, the answer is, not necessarily.

[Kalai and Lehrer (1993)] was not the only work at that time to explore questions of convergence under rational learning; in fact, other influential works that both support it and contrast with it were carried out around the same time

[*]Nuffield College and Department of Economics, Oxford University, john.calculus@gmail.com

[†]The author expresses thanks to Peyton Young for his helpful remarks and for bringing to the author's attention the conjecture this paper addresses, and to Ron Peretz for his help with the probabilistic tools this paper utilizes.

and in the following years. [Jordan (1991)], for example, shows that under appropriate assumptions, rational learning converges to the set of stage game Nash equilibria. However, [Kalai and Lehrer (1993)] struck on a property of learning discussed very little at the time: the convergence of the agents' strategies to (approximate) equilibria of the *repeated game*, not of the strategic game that is being repeated. While their assumptions that the players' beliefs all contain a grain of truth[1] have been scrutinised as both an over-demanding coordination requirement (e.g., [Miller and Sanchirico (1999)]) and a highly non-generic condition (e.g., [Miller and Sanchirico (1997)]), the resulting body of literature extending, discussing, and contrasting [Kalai and Lehrer (1993)] "is, in many respects, a natural successor to the earlier literature on learning rational expectations" as "both literatures address the question of whether decision-makers can, through repeated experience, learn to make optimal or equilibrium decisions" ([Jordan (1993)]).

A primary and motivating example in which such learning occurs naturally is the class of *Bayesian games*, in particular when the preferences of the agents – that is, their *types* – are not known publicly but others do have beliefs about them, which are updated at each stage. Such interactions occur naturally, for example, in sequential auctions, where the private values of the object being sold are not commonly known; however, agents learn more about the others' preferences as time goes on and bids are observed, e.g. [Jeitschko (1998)]. Sequential bargaining with incomplete information, wars of attrition, and repeated duopolistic competition when others' costs are uncertain all naturally fit under this umbrella framework as well.

Two main veins of subsequent work exist. One direction generalises the results of [Kalai and Lehrer (1993)], some works by weakening the absolute continuity assumptions on the beliefs, as in [Sandroni (1998)] and [Norman (2012)], others by weakening assumptions on the players' knowledge, as in [Kalai and Lehrer (1995)], [Jordan (1995)], and [Nyarko (1998)], and still other variations, e.g., [Gilli (2001)] and the references there. Another direction, however, was to point out the limitations of the assumptions and results, as in the papers [Lehrer and Smorodinsky (1996)], [Lehrer and Smorodinsky (1997)], [Miller and Sanchirico (1997)] and [Miller and Sanchirico (1999)], [Nachbar (1997)], and [Foster and Young (2001)].

The contribution of this paper is to answer a long-standing open question raised in [Kalai and Lehrer (1993), Sec 7.1]. The classical result of that paper ensures convergence of the play to the set of *approximate* equilibria (i.e., ε-equilibria) of the repeated game. (One cannot in general expect convergence to a specific equilibrium or approximate equilibrium, as players may, for example, rotate among different equilibria.) The authors raise, but leave open, the question of whether play must converge to the set of *exact* equilibria. In this paper we show by example that this need not be the case. Furthermore, not only

[1] Or somewhat weaker absolute continuity conditions.

does convergence fail to occur, but the play induced by any Nash equilibrium remains far – in fact, mutually singular – from actual play.

[Kalai and Lehrer (1993), Sec 6] address in particular the question of rational learning in certain Bayesian games, which, as mentioned previously, are a primary and general class of such interactions. In these games, each player is privately assigned a type (types are chosen independently) and a player's payoff may depend on his own type. Players can then condition their actions both on the public play so far and on their type. [Kalai and Lehrer (1993), Sec 6] show that as time goes by, play must converge to the set of the *approximate* equilibria of the true game, or informally (p. 1038 there), "even if the players do not learn the identity of the payoff matrices actually played, they eventually play almost as ε-Nash players who do know the identity of the payoff matrices." We also show in our paper that this result cannot be strengthened to deduce convergence to the set of plays induced by equilibria of the true game.

Our results are also interesting in light of the results of [Kalai and Lehrer (1993), Sec 7.1]. There, an additional result is given which shows that rational learning does approach the exact equilibria provided one uses a much weaker notion of closeness, one which only guarantees closeness over a fixed finite horizon. Hence, combined with the result on merging of beliefs of [Blackwell and Dubins (1962)], our work shows that although agents will make predictions which not only merge towards each other but resemble Nash equilibria *in the short run*, their beliefs may *not* merge towards the predictions made *in the long run* by Nash equilibria. We elaborate on the results of [Kalai and Lehrer (1993), Sec 7.1], and contrast it with the stronger notion of learning, in Section 9.

The rational learning model is presented formally in Section 2. A brief informal overview of the construction is given in Section 3. The stage game of our example is presented in Section 4, while the strategies and beliefs are given in Section 5. In order to prove that convergence to Nash equilibrium does not occur, Section 6 contains a preliminary result, while the proof itself is given in Section 7. Section 8 presents the Bayesian game. Section 9 compares our result to the short-term merging to equilibria discussed in [Kalai and Lehrer (1993), Sec 7.1]. Section 10 presents a slightly different example, with its own virtues and disadvantages. Some probabilistic tools appear in Appendix A, while proofs from Section 6 appear in Appendix B.

2 The Rational Learning Model

Let $\mathcal{P}$ be a finite set of players with finite action spaces $(A^k)_{k\in\mathcal{P}}$, and let G be a strategic game on this set of players with payoff functions $r = (r^k)_{k\in\mathcal{P}}$; hence $r^k : \prod_{k\in\mathcal{P}} A^k \to \mathbb{R}$ for each $k \in \mathcal{P}$. For $T = 0, 1, 2, \ldots, \infty$, let $H_{T+1} =$

$(\prod_{k\in\mathcal{P}} A^k)^T$ be the collection of histories of the T-stage repeated game[2] (with $H_1 = \{\emptyset\}$). Denote $H_* = \underset{t<\infty}{\cup} H_t$.

A *behavioural strategy* for Player $k \in \mathcal{P}$ is a mapping[3] $\sigma^k : H_* \to \Delta(A^k)$. A profile of strategies $\sigma = (\sigma^k)_{k\in\mathcal{P}}$ induces a measure P_σ (and associated expectation operator E_σ) in the T-stage repeated games in the usual way, i.e., on H_{T+1}, for each $T = 0, 1, 2, \ldots, \infty$, defined by $P_\sigma(a_1, \ldots, a_t) = \prod_{s<t} \sigma(a_1, \ldots, a_{s-1})[a_s]$; P_σ extends naturally to H_∞. The payoff in the infinitely repeated game G^∞ is given by

$$\overline{r}(a_1, a_2, \ldots) = \sum_{t=1}^{\infty} \beta^{t-1} r(a_t)$$

where $0 < \beta < 1$ is a fixed discount factor. For $\varepsilon \geq 0$, a strategy profile σ of G^∞ is an ε-*equilibrium* (or just *equilibrium* when $\varepsilon = 0$) if for each player $k \in \mathcal{P}$ and each strategy τ of Player k,

$$E_\sigma[\overline{r}^k] + \varepsilon \geq E_{(\tau,(\sigma^j)_{j\neq k})}[\overline{r}^k]$$

For each $h = (a_1, \ldots, a_T) \in H_*$, for $t \leq T$ we denote $h|_t = (a_1, \ldots, a_{t-1})$, and for $k \in \mathcal{P}$ and strategy σ^k of Player k, let σ^k_h be the strategy defined by $\sigma^k_h(a'_1, \ldots, a'_t) = \sigma^k(a_1, \ldots, a_T, a'_1, \ldots, a'_t)$, and similarly for profiles of strategies. To recall the results of [Kalai and Lehrer (1993)] and to state our results clearly, we introduce the following notion:

Definition 2.1. *Let $(\tau^{j,k})_{j,k\in\mathcal{P}}$ be a $\mathcal{P} \times \mathcal{P}$ collection of strategies, where $\tau^{j,k}$ is a strategy of Player k (interpreted as the belief of Player j about Player k's strategy). We say that in $(\tau^{j,k})_{j,k\in\mathcal{P}}$ each player is best-replying to his beliefs if for each $j \in \mathcal{P}$, $\tau^{j,j}$ is a best reply to $(\tau^{j,k})_{k\neq j}$, i.e., for each strategy π of Player j,*

$$E_{(\tau^{j,k})_{k\in\mathcal{P}}}[\overline{r}^j] \geq E_{(\pi,(\tau^{j,k})_{k\neq j})}[\overline{r}^j]$$

Given such $\tau = (\tau^{j,k})_{j,k\in\mathcal{P}}$, by P_τ we mean the distribution induced by the diagonal $(\tau^{j,j})_{j\in\mathcal{P}}$, interpreted as the strategies which are *actually* played (since each player knows his own strategy).

For a measurable space $(\Omega, \mathcal{B})$, the set of probability measures on Ω is denoted $\Delta(\Omega)$, and for each $\mu, \nu \in \Delta(\Omega)$, the *total variation distance* between μ, ν is:

$$||\mu - \nu|| = 2 \cdot \sup_{A\in\mathcal{B}} |\mu(A) - \nu(A)|$$

μ is *absolutely continuous w.r.t.* ν, denoted $\mu \ll \nu$, if for all $A \in \mathcal{B}$, $\nu(A) = 0$ implies $\mu(A) = 0$. μ and ν are *mutually singular*, denoted $\mu \perp \nu$, if there is $A \in \mathcal{B}$ such that $\mu(A) = \nu(\Omega \backslash A) = 1$. Observe that $\mu \perp \nu$ implies $||\mu - \nu|| = 2$.

[2] It is convenient to denote the set of histories of the T-stage game as H_{T+1}, since we will view this set as the space of plays *preceding* stage $T + 1$.

[3] For a set X, $\Delta(X)$ denotes the space of probability measures on X.

Definition 2.2. *Let P, Q be probability measures (on the same space); then Q contains a grain of truth of P if $P \ll Q$ and the Radon-Nikodym derivative $\frac{dP}{dQ}$ is bounded; equivalently, if for some $0 < \lambda < 1$ and some probability measure P', $Q = \lambda P + (1-\lambda)P'$.*

If Ω is any set and $f : \Omega \to \mathbb{R}^N$, denote $||f||_\infty = \sup_{\omega \in \Omega, n \leq N} |f^n(\omega)|$. We use this notation also for $p, q \in \Delta(\Omega)$ for finite Ω: $||p - q||_\infty = \max_{\omega \in \Omega} |p[\omega] - q[\omega]|$.[4]

The following is the main result of [Kalai and Lehrer (1993)]:

Theorem 2.3. *Let $\tau = (\tau^{j,k})_{j,k \in \mathcal{P}}$ be such that everyone is best-replying to their beliefs. Denote by $\tilde{\tau}^j = (\tau^{j,k})_{k \in \mathcal{P}}$ the beliefs of Player j. Assume that for each $j \in \mathcal{P}$, $P_\tau \ll P_{\tilde{\tau}^j}$. Then for P_τ-a.e. $h \in H_\infty$ and any $\varepsilon > 0$, there exists $T \in \mathbb{N}$ such that for each $t \geq T$, there is an ε-equilibrium σ of G^∞ such that $||P_{\tau_{h|_t}} - P_\sigma|| < \varepsilon$.*

It should be emphasised that this result does not say that there exists any particular ε-equilibrium σ such that from *any* late enough time period t, the play induced in the game after t stages is close to the play induced by σ.[5] The result says that the distribution of play induced after enough time is *close to the set* of possible distributions of play induced by all ε-equilibria. That is, at some periods, the induced play may be close to the play induced by one ε-equilibrium, but at other periods, close to the play induced by another.

In Section 7.1 of [Kalai and Lehrer (1993)], the following question is posed: Can Theorem 2.3 be strengthened to require that the process converges to the set of *exact equilibria* (that is, 0-equilibria) of G^∞, and not just the set of ε-equilibria? This question has remained open until now. The purpose of this paper is to present a counter-example showing that we may not, in general, be able to guarantee that play of the game will eventually resemble the play of an *exact* equilibrium. To be more precise:

Result 2.4. *We construct a game G and $(\tau^{j,k})_{j,k \in \mathcal{P}}$ in which everyone is best-replying to their beliefs, such that for each $j \in \mathcal{P}$, $P_\tau \ll P_{\tilde{\tau}^j}$ – in fact, the beliefs contain a grain of truth – and such that for P_τ-a.e. $h \in H_\infty$, for each $t \in \mathbb{N}$, and each equilibrium σ of G^∞, $P_{\tau_{h|_t}} \perp P_\sigma$.*

A particular type of beliefs arises from *repeated Bayesian games* (a.k.a. *games of incomplete information*). Each player has a discrete type space, $(T^k)_{k \in \mathcal{P}}$. At the beginning of play, types are chosen by Nature before play begins independently, i.e., via a commonly known product distribution $\mu = \prod_{k \in \mathcal{P}} \mu_k$. Each player is informed of his own type, hence a strategy for Player k is a

[4] Hence, on $\Delta(\Omega)$ for finite Ω, we have both the supremum norm $|| \cdot ||_\infty$ and the total variation norm $|| \cdot ||$.

[5] Indeed, if the stage game possesses multiple equilibria and if the strategies alternate between these equilibria, we will would never get such convergence.

mapping $\eta^k : T^k \times H_* \to \Delta(A^k)$, and a profile $\eta = (\eta^k)_{k\in\mathcal{P}}$ of such strategies together with the prior μ induces a measure P_η on $\prod_{k\in\mathcal{P}} T^k \times H_\infty$. The payoff of Player k may include dependence on his type (and only his type, not the types of the others), and hence is a (bounded) function $r^k : T^k \times \prod_{k\in\mathcal{P}} A^k \to \mathbb{R}$. Let $G^\infty(\mu)$ denote this game of incomplete information, and for each profile of types $(t^k)_{k\in\mathcal{P}} \in \prod_{k\in\mathcal{P}} T^k$, let $G((t^k)_{k\in\mathcal{P}})$ be the strategic game resulting from this selection of types (i.e., the one-shot game in which the types $(t^k)_{k\in\mathcal{P}}$ are selected and made public), and let $G^\infty((t^k)_{k\in\mathcal{P}})$ be the game in which the strategic-form game $G((t^k)_{k\in\mathcal{P}})$ is repeated infinitely many times.[6] Defining the payoffs also to be the discounted sum of the stream of the stage payoffs, Bayesian equilibrium is defined in the same way, that is, no player can increase his expected payoff by deviating.[7]

Theorem 2.5. *[[Kalai and Lehrer (1993)] , Sec. 6] Let $\eta = (\eta^j)_{j\in\mathcal{P}}$ be a Bayesian equilibrium of $G^\infty(\mu)$. Then for each $\varepsilon > 0$, for μ-a.e. choice of types $(t^k)_{k\in\mathcal{P}} \in \prod_{k\in\mathcal{P}} T^k$ of Nature, for $P_\eta(\cdot \mid (t^k)_{k\in\mathcal{P}})$-a.e. play $h \in H_\infty$, there is a time T such that for each $t \geq T$, there is an ε-equilibrium $\sigma = (\sigma^j)_{j\in\mathcal{P}}$ of $G^\infty((t^k)_{k\in\mathcal{P}})$, such that[8] $||P_{\eta_{h|_t}}(\cdot \mid (t^k)_{k\in\mathcal{P}}) - P_\sigma|| < \varepsilon$.*

In this paper:

Result 2.6. *We construct a repeated Bayesian game $G^\infty(\mu)$ with finite type spaces $(T^k)_{k\in\mathcal{P}}$ and a Bayesian equilibrium $(\eta^k)_{k\in\mathcal{P}}$ of $G^\infty(\mu)$, such that for some choice of types $(t^k)_{k\in\mathcal{P}} \in \prod_{k\in\mathcal{P}} T^k$ by Nature (of positive μ-measure) and $P_\eta(\cdot \mid (t^k)_{k\in\mathcal{P}})$-a.e. play $h \in H_\infty$, for each $t \in \mathbb{N}$, and each equilibrium σ of $G^\infty((t^k)_{k\in\mathcal{P}})$, $P_{\eta_{h|_t}}(\cdot \mid (t^k)_{k\in\mathcal{P}}) \perp P_\sigma$.*

We raise a question worthy of further investigation but that we shall not attempt to answer here. In the examples presented in this paper, the payoffs will be seen to be highly non-generic, and their success in evading rational learning converging to equilibria is dependent precisely on how they are defined. Hence, it is natural to inquire: does convergence to the set of equilibria occur for generic games, i.e., that the set of stage games for which this strengthened convergence does not hold is small in an appropriate sense (first category, measure zero, etc.)? Another question worth investigating is whether convergence to the set of equilibria holds for two-player games.

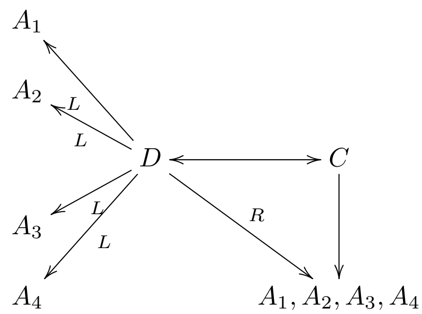
[6] To clarify: The types are chosen once and become common knowledge at the beginning of play, and then the game is infinitely repeated with those fixed types.

[7] Since the type space is discrete, we needn't differentiate between ex-ante deviations, i.e., deviations proceeding the selections of types, or ex-post deviations, i.e., deviations following the selection of types. If the type space were more general, technical measurability issues would arise. However, Theorem 2.5 does not remain correct in such a general framework anyway; see [Lehrer and Smorodinsky (1997)].

[8] $\eta_{h|_t}$ denotes, as earlier, the strategy in the Bayesian game induced by η following play of $h|_t$. Strictly speaking, the play following $h|_t$ is not a sub-game as in itself since it does not include the previous private announcements of the types.

3 Informal Construction Overview

- The game has six players, $\mathcal{P} = \{A_1, A_2, A_3, A_4, C, D\}$. Each player has two actions, L and R.
- C and D play a coordination game between them. When D plays L, everyone else is incentivised to play L as well. When D plays R, the players A_1, A_2, A_3, A_4 play a sort of anti-coordination game among themselves, which also depends on C. We will focus on a mixed equilibrium of this game in which $A_1, \ldots, A_4$ each mix $\frac{1}{2}$-$\frac{1}{2}$. The equilibrium in which all play L is far preferred by all players.
- The influence of the players on each other can be viewed via the following graph:



- When C and D coordinate on (R, R) for the n-th time, players A_1, A_2, A_3, A_4 will incorrectly put some small probability δ_n^3 on C playing L.[9] As a result, equilibria for A_1, A_2, A_3, A_4 deviate slightly from $\frac{1}{2}$-$\frac{1}{2}$. We will concentrate on the deviation in which each plays L with probability $\frac{1}{2} + \delta_n$.
- Although the beliefs about C's action contain only a small error, on the order of δ_n^3 the n-th time C and D coordinate on R, this leads to a much larger – order of δ_n – deviation in equilibrium strategies for A_1, A_2, A_3, A_4. Although (δ_n) will converge to zero quickly enough to guarantee that the beliefs contain a grain of truth, they will shrink slowly enough so that if an equilibrium *of the stage game* is used *at each stage*, the play of the game must be far (in fact, mutually singular) from the true play.
- Thus, to avoid equilibria in the repeated game in which players do not play equilibria in the stage game – by implementing intertemporal dependence and coordination – we have in our strategies and beliefs in between such stages at which C and D coordinate on (R, R) long periods in which they coordinate on (L, L), hence incentivising A_1, A_2, A_3, A_4 to play L as well. Hence, at those stages in which C and D coordinate on (R, R), play must

[9] This can be justified, as we will do later in the framework of Bayesian games, by believing that with small probability C is actually indifferent between his actions and that he puts positive probability on both.

be very close to an equilibrium of the stage game, since all know that the next time any player has an incentive to play anything other than L – which, recall, is a much-preferred outcome for all – will be so far into the future and hence so discounted that no player has any hope that a deviation could lead to a future profit.

4 The Stage Game

The stage game G has six players, which we denote $\mathcal{P} = \{A_1, A_2, A_3, A_4, C, D\}$. Each player has action set $A = \{L, R\}$. Let x be a mixed action profile. For $k \in \mathcal{P}$, we will write x^k instead of $x^k[L]$ (the probability that k plays L). For $j = 1, 2, 3, 4$, denote $\Delta^{-j}(x) = \prod_{k \neq j}(x^{A_k} - \frac{1}{2})$.

The payoff to Players C, D depends only on each other, and is a coordination game given by:

$r^{C,D}(\cdot, a^{-\{C,D\}}) =$

$C \backslash D$	L	R
L	$2, 2$	$0, 0$
R	$0, 0$	$1, 1$

For $j = 1, 2, 3, 4$, the payoff to A_j is given by:

$r^{A_j}(\cdot, x^{-A_j}) =$

	$a^D = L$	$a^D = R$
L	2	$x^C - \Delta^{-j}(x)$
R	0	0

This gives a well-defined payoff since the payoff for A_j depends in a multi-linear way on the others' actions.

Lemma 4.1. *For each $0 < \delta < 1$, there exists $\eta > 0$ such that there is no η-equilibrium z of G satisfying:*

- $z^C = z^D = R$.
- *For* $j = 1, 2, 3, 4$, $\frac{1}{2} + \frac{1}{2}\delta \leq z^{A_j}$.

Proof. Fix $0 < \delta < 1$. Suppose for each $\eta > 0$, there was an η-equilibrium z_η satisfying these conditions. Taking a limit of a subnet of $(z_\eta)_{\eta > 0}$ would give an equilibrium z of G satisfying these properties, and in particular, since $z^C = R$, for all $j = 1, 2, 3, 4$, $z^C - \Delta^{-j}(z) \leq -\frac{1}{8}\delta^3 < 0$; hence, since $z^D = R$, we would have $z^{A_j} = R$ for all $j = 1, 2, 3, 4$, a contradiction since $z^{A_j} \geq \frac{1}{2} + \frac{1}{2}\delta$. □

It is immediate to verify:

Lemma 4.2. *The profile $\overline{L}$ with $\overline{L}^k = L$ for each $k \in \mathcal{P}$ is an equilibrium of G.*

5 The Strategies and Beliefs

Let $(\delta_n)_{n=1}^{\infty}$ be any positive sequence with

$$\sup_n \delta_n < \min[\frac{1}{4}, \varepsilon, \zeta^{-1}], \quad \sum_{n=1}^{\infty} \delta_n^3 < \infty, \quad \sum_{n=1}^{\infty} \delta_n^2 = \infty \tag{5.1}$$

where $\varepsilon, \zeta > 0$ are specified later in Proposition 6.1, and

$$L_n = T_0(\delta_n) \in \mathbb{N} \tag{5.2}$$

with $T_0(\cdot)$ also defined later in Proposition 6.1. Denote $S_n = 1 + \sum_{k<n} L_n$ (note that $S_1 = 1$). We define the following array of strategies $(\tau^{j,k})_{j,k\in\mathcal{P}}$; they are history-independent and hence we will not reference the history explicitly in the definition, i.e., we will denote $\tau_t^{j,k}$ instead of $\tau^{j,k}(h|_t)$. Recall also that we denote τ^k instead of $\tau^{k,k}$.

- $$\tau_t^C = \tau_t^D = \begin{cases} R & \text{if } \exists n \in \mathbb{N}, t = S_n \\ L & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

- For each $j = 1, 2, 3, 4$,

$$\tau_t^{A_j} = \begin{cases} \frac{1}{2} + \delta_n & \text{if } t = S_n \\ L & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

- All players have correct beliefs about A_1, A_2, A_3, A_4, D: Formally, $\tau^{k,A_j} = \tau^{A_j}$, $\tau^{k,D} = \tau^D$, for $k \in \mathcal{P}$ and $j = 1, 2, 3, 4$.

- However, for all $k \in \mathcal{P}$, $k \neq C$, the beliefs of Player k about C are

$$\tau^{k,C} = \begin{cases} \delta_n^3 & \text{if } t = S_n \\ L & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

Recall that we denote the beliefs of Player $j \in \mathcal{P}$ by $\tilde{\tau}^j := (\tau^{j,k})_{k\in\mathcal{P}}$, and similarly we denote $\tilde{\tau}_t^j := (\tau_t^{j,k})_{k\in\mathcal{P}}$.

Lemma 5.1. *In $\tau = (\tau^{j,k})_{j,k\in\mathcal{P}}$, each player is best-replying to his beliefs.*

Proof. Because of the history independence, it's enough to check that for each $t \in \mathbb{N}$, each $k \in \mathcal{P}$, τ_t^k is a best-reply to his beliefs $\tilde{\tau}_t^k$ in the stage game G. If $\forall n, t \neq S_n$, then τ_t is just the equilibrium $\overline{L}$ from Lemma 4.2 and all players have the correct beliefs of the other players' action at stage t, i.e, $\tau_t^{j,k} = \tau_t^k$. Suppose $t = S_n$. We contend that the (A_j) are all indifferent; indeed, $r^{A_j}(R, \cdot) \equiv 0$, and since $\tau_{S_n}^{A_j,D} = \tau_{S_n}^D = R$,

$$r^{A_j}(L, (\tau_{S_n}^{A_j,k})_{k\neq A_j}) = \tau_{S_n}^{A_j,C} - \Delta^{-j}(\tilde{\tau}_{S_n}^{A_j}) = \delta_n^3 - ((\frac{1}{2} + \delta_n) - \frac{1}{2})^3 = 0$$

Furthermore, since $\tau_t^{C,D} = \tau_t^D = R$, C prefers R; and D prefers R since $\tau_t^{D,C} = \delta_n^3 < (\frac{1}{4})^2 < \frac{1}{3}$. □

Lemma 5.2. *For each player $k \in \mathcal{P}$, $P_\tau \ll P_{\tilde\tau^k}$, and in fact, $P_{\tilde\tau^k}$ contains a grain of truth of P_τ.*

Proof. C has correct beliefs about all players; i.e., $\tilde\tau^C = \tau$. If $k \neq C$, then denoting

$$u_n^C = \delta_n^3, u^D = R,\ u_n^{A_j} = \frac{1}{2} + \delta_n \quad j = 1, 2, 3, 4$$

one verifies that the beliefs of $A_1, \ldots, A_4, D$ satisfy

$$\tilde\tau_t^{A_j} = \tilde\tau_t^D = \begin{cases} u_n & \text{if } t = S_n \\ \overline{L} & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

where $\overline{L}$ is the equilibrium of the stage game given by $\overline{L}^k = L$ for all $k \in \mathcal{P}$. Let v_n denote the profile

$$v_n^C = v_n^D = R,\ v_n^{A_j} = u_n^{A_j} = \frac{1}{2} + \delta_n \quad j = 1, 2, 3, 4 \tag{5.3}$$

Then, similarly, the players' strategies are

$$\tau_t = \begin{cases} v_n & \text{if } t = S_n \\ \overline{L} & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

Since $\forall a \in \{L, R\}^{\mathcal{P}}$, $u_n[a] = 0$ implies $v_n[a] = 0$, by Proposition 11.1 it suffices to check that

$$\sum_{n=1}^{\infty} ||u_n - v_n||_\infty < \infty$$

and that there is some $\alpha > 0$ such that for all $t \in \mathbb{N}$ and all profile $a \in \{L, R\}^{\mathcal{P}}$, if $v_n[a] > 0$ then $u_n[a] \geq \alpha$. For the latter claim, take[10] $\alpha = \frac{1}{2}(\frac{1}{4})^4$. For the former, observe that $||u_n - v_n||_\infty \leq 2\sum_{k\in\mathcal{P}} |u_n^k - v_n^k| \leq 2|\mathcal{P}|\delta_n^3$, and by (5.1), $\sum_{n=1}^{\infty} \delta_n^3 < \infty$. □

6 Equilibrium Analysis

For $T \in \mathbb{N}$, denote

$$H_{T+1}^\alpha := \{h = (a_1, \ldots, a_T) \in H_{T+1} \mid a_1^C = a_1^D = R \text{ and } \forall k \in \mathcal{P}, 2 \leq t \leq T, a_t^k = L\} \tag{6.1}$$

i.e., those histories of length T in which Players C, D plays R in the first round, and after the first round, everyone plays only L. Note that H_{T+1}^α is finite, $|H_{T+1}^\alpha| = 2^4$. The following is the main proposition that we will require:

[10] Since $v_n[a] > 0$ implies $a^C = a^D = R$, the definition of u_n shows that $v_n[a] > 0 \to u_n[a] \geq (1 - \delta_n^3)(\frac{1}{2} - \delta_n)^4$. Recall then that $\delta_n < \frac{1}{4}$.

Proposition 6.1. *Let $\zeta = 4^4$, let $0 < \delta < \frac{1}{2}$, let $\varepsilon > 0$ satisfy*

$$\zeta\varepsilon(2 + ||r||_\infty)\frac{1}{1-\beta} < \frac{1}{4} \tag{6.2}$$

and let $T_0 = 2 \cdot T_1$ for T_1 satisfying

$$\beta^{T_1}\frac{||r||_\infty}{1-\beta} \leq \min[\frac{\eta}{2}, \frac{1}{8}] \tag{6.3}$$

where $\eta > 0$ corresponds to δ as in Lemma 4.1. Let $T \geq T_0$. Then there does not exist an equilibrium profile σ which satisfies:

- *For all $j = 1, 2, 3, 4$, $\frac{1}{2} + \frac{1}{2}\delta \leq \sigma^{A_j}(\emptyset)$.*[11]
- *$P_\sigma(H^\alpha_{T+1}) > 1 - \varepsilon$ and $P_\sigma(h) > \frac{1}{2}\zeta^{-1}$ for each $h \in H^\alpha_{T+1}$.*

Suppose σ were such an equilibrium. The following lemmas are proved in Appendix B:

Lemma 6.2. *For all $2 \leq t \leq T - T_1$, each $h \in H^\alpha_t$, and each player $k \in \mathcal{P}$, $\sigma^k(h) = L$.*

The idea is that if enough stages (at least T_1) are remaining, each player knows that that since $P_\sigma(H^\alpha_{T+1}) > 1 - \varepsilon$, if he doesn't deviate he likely faces only L's being played for a long time - which is a desired outcome for him, and hence he will not want to ruin the possibility by deviating.

Lemma 6.3. *$\sigma^C(\emptyset) = \sigma^D(\emptyset) = R$.*

The intuition is the same: Players C, D do not want to ruin the prospect of a long stream of L's being played by deviating out of H^α_{T+1}.

Lemma 6.4. *$\sigma(\emptyset)$ is a η-equilibrium of the stage game G, where η corresponds to δ as in Lemma 4.1.*

The intuition is that, since C, D are coordinating on R in the initial stage by Lemma 6.3, regardless of what $A_1, \ldots, A_4$ play in the first stage σ will (by Lemma 6.2) follow with a long stream of L's; hence, the action profile chosen by $A_1, \ldots, A_4$ will have very little affect on future payoffs, and hence should be close to an approximate equilibrium response of the stage game.

(Proof of Proposition 6.1). By Lemma 6.4, $z := \sigma(\emptyset)$ is an η-equilibrium of the stage game. Hence, we cannot have both the conditions listed in Lemma 4.1 holding. However, Lemma 6.3 shows indeed that $z^C = z^D = R$, while by assumption, $z^{A_j} \geq \frac{1}{2} + \frac{1}{2}\delta$ for each $j = 1, 2, 3, 4$, a contradiction. □

[11] Recall that $\emptyset$ denotes the empty history.

7 Proof of No Convergence to Equilibria

Recall the array $\tau = (\tau^{j,k})_{j,k\in\mathcal{P}}$ defined in Section 5.

Proposition 7.1. *For any equilibrium σ of G^∞, $P_\tau \perp P_\sigma$.*

Proof. For each $n \in \mathbb{N}$, let $X_n = (\{L,R\}^{\mathcal{P}})^{L_n} = H_{L_n+1}$ be the possible sequences of play in the L_n-repeated game. Then clearly $H_\infty = X := \prod_{n\in\mathbb{N}} X_n$, while $H_{S_n} = \overline{X}_n := \prod_{k<n} X_k$, where recall that $S_n = 1 + \sum_{k<n} L_k$. (Essentially, we've partitioned the stages into blocks $X_1, X_2, \ldots$ of lengths $L_1, L_2, \ldots$) We use the following notations:

For a strategy profile τ and its induced measure $P_\tau \in \Delta(X)$, let $(P_\tau)_n$ denote the marginal of P_τ on $\overline{X}_{n+1} = H_{S_{n+1}} = \prod_{k\le n} X_k$, and let $(P_\tau)_n[\cdot|\cdot]$ be the conditional on X_n w.r.t. $\overline{X}_n$: I.e., $(P_\tau)_n[\cdot \mid \overline{x}_n]$ is the distribution on X_n given $\overline{x}_n = (x_1, \ldots, x_{n-1}) \in \overline{X}_n$; if $P_\tau(\overline{x}_n) = 0$, then these conditional distributions are arbitrary, and any specifications of it are called a *version* of $(P_\tau)_n[\cdot \mid \cdot]$. (Referring back to our game, $(P_\tau)_n$ is the distribution on the L_n-block of the repeated game, given the play of the proceeding blocks of the sizes $L_1, \ldots, L_{n-1}$.)

Now in our case, for our specific strategy profile τ we have the version of $(P_\tau)_n[\cdot \mid \cdot]$ given by $(P_\tau)_n[\cdot \mid \cdot] \equiv v_n \otimes_{k=2}^{L_n} \overline{L}$, with v_n defined in (5.3), and $\overline{L}^k = L$ for all $k \in \mathcal{P}$.

On the other hand, for $\overline{x}_n \in \overline{X}_n = \prod_{k<n} X_k = H_{S_n}$, letting σ be an equilibrium of G^∞, we have the version of $(P_\sigma)_n[\cdot \mid \overline{x}_n]$ given by $(P_\sigma)_n[x_n \mid \overline{x}_n] = P_{\sigma_{\overline{x}_n}}(x_n)$ when $P_\sigma(\overline{x}_n) > 0$ (and arbitrary otherwise).

For all $\overline{x}_n \in \overline{X}_n$ with $P_\sigma(\overline{x}_n) > 0$, $\sigma_{\overline{x}_n}$ is an equilibrium as well.[12] Now, fix such $\overline{x}_n \in \overline{X}_n$ with $P_\sigma(\overline{x}_n) > 0$; for brevity denote $\hat\sigma$ instead of $\sigma_{\overline{x}_n}$ and $\hat\tau$ instead of $\tau_{\overline{x}_n}$. Applying Proposition 6.1 (recall $L_n = T_0(\delta_n)$ by (5.2)) implies that one of the following holds:

(i) For some $j = 1,2,3,4$, $\hat\sigma^{A_j}(\emptyset) < \frac{1}{2} + \frac{1}{2}\delta_n$.

(ii) $P_{\hat\sigma}(H^\alpha_{T+1}) \le 1 - \varepsilon$, where H^α_{T+1} is defined just before Proposition 6.1.

(iii) $P_{\hat\sigma}(h) \le \frac{1}{2}\zeta^{-1}$ for some $h \in H^\alpha_{T+1}$.

Going over these case-by-case:

(i) In this case,

$$|\hat\sigma^{A_j}(\emptyset) - \hat\tau^{A_j}(\emptyset)| = |\hat\sigma^{A_j}(\emptyset) - (\frac{1}{2} + \delta_n)| \ge \frac{\delta_n}{2}$$

so $||(P_\tau)_n[\cdot \mid \overline{x}_n] - (P_\sigma)_n[\cdot \mid \overline{x}_n]|| \ge \delta_n$.

[12] Indeed, if σ is an equilibrium and $h \in \cup_{t=1}^\infty H_t$ with $P_\sigma(h) > 0$, then σ_h is also an equilibrium.

(ii) $P_{\hat{\sigma}}(H^{\alpha}_{L_n+1}) \leq 1-\varepsilon$ while $P_{\hat{\tau}}(H^{\alpha}_{L_n+1}) = 1$, hence $||(P_\tau)_n[\cdot \mid \overline{x}_n] - (P_\sigma)_n[\cdot \mid \overline{x}_n]|| \geq 2\varepsilon \geq \delta_n$ by (5.1).

(iii) $P_{\hat{\sigma}}(h) \leq \frac{1}{2}\zeta^{-1}$, while $P_{\hat{\tau}}(h) \geq \zeta^{-1}$; hence $||(P_\tau)_n[\cdot \mid \overline{x}_n] - (P_\sigma)_n[\cdot \mid \overline{x}_n]|| \geq \zeta^{-1} \geq \delta_n$ by (5.1).

Hence, $||(P_\tau)_n[\cdot \mid \overline{x}_n] - (P_\sigma)_n[\cdot \mid \overline{x}_n]|| \geq \delta_n$. This was for any $\overline{x}_n \in \cup_n \overline{X}_n$ with $P_\sigma(\overline{x}_n) > 0$. Hence, by Theorem 11.4, since $\sum_{n=1}^{\infty} \delta_n^2 = \infty$, $P_\tau \perp P_\sigma$. $\square$

Corollary 7.2. *For P_τ-a.e. $h \in H_\infty$, for all $t \in \mathbb{N}$, $P_{\tau_{h|_t}} \perp P_\sigma$ for any equilibrium σ of G^∞.*

Proof. Let $t \in \mathbb{N}$ and $h^* \in H_t$, and fix an equilibrium σ of G^∞. Denote $\tau' = \tau_{h^*}$. We will show that $P_{\tau'} \perp P_\sigma$.

First take the case that $t = S_n$ for some n. In this case, one simply observes that τ' is induced by the sequence $(L_n, L_{n+1}, \ldots)$ in the same way that τ is induced by the sequence $(L_1, L_2, \ldots)$, and applies Proposition 7.1 (since the sequence $(\delta_n, \delta_{n+1}, \ldots)$ also satisfies (5.1), and $L_n = T_0(\delta_n)$).

Suppose now $S_n < t < S_{n+1}$. Denote $T = S_{n+1} - t$. Fix some $h' \in H_{T+1}$ (hence, $(h^*, h') \in H_{S_{n+1}}$) such that $P_{\tau'}(h') > 0$ and also $P_\sigma(h') > 0$. Let $\tau'' = \tau'_{h'}$, $\sigma' = \sigma_{h'}$; then σ' is an equilibrium of G^∞ since $P_\sigma(h') > 0$, and τ'' is induced by the sequence $(L_{n+1}, L_{n+2}, \ldots)$ in the same way that τ is induced by the sequence $(L_1, L_2, \ldots)$. Hence, like the case above, $P_{\tau''} \perp P_{\sigma'}$, i.e. $P_{\tau'_{h'}} \perp P_{\sigma_{h'}}$, and therefore $P_{\tau'}(\cdot \mid h') \perp P_\sigma(\cdot \mid h')$. To sum up, for each $h' \in H_{T+1}$ such that $P_{\tau'}(h') > 0$ and also $P_\sigma(h') > 0$, we have $P_{\tau'}(\cdot \mid h') \perp P_\sigma(\cdot \mid h')$. Hence, from Proposition 11.5, $P_{\tau'} \perp P_\sigma$, as required. $\square$

8 The Bayesian Game

Relying on the payoffs r defined in Section 4 and the strategies and beliefs $\tau = (\tau^{j,k})_{j,k \in \mathcal{P}}$ defined in Section 5, we now define a Bayesian game:

- First, let $(\delta_n)_{n=1}^{\infty}$ be a positive sequence satisfying, in addition to (5.1), the condition $\delta_n < \frac{1}{2\sqrt[3]{3}}$, and

$$\prod_{k=1}^{\infty} \frac{1}{1-\delta_k^3} \leq \frac{4}{3} \tag{8.1}$$

  (E.g., take $\delta_n = \frac{1}{\sqrt{n+M}}$, for large enough $M > 0$.)

- The set of players $\mathcal{P} = \{C, D, A_1, A_2, A_3, A_4\}$ is the same as before, each with actions $\{L, R\}$.

- Player C has two types, which we denote $t^C = \uparrow$ or $t^C = \downarrow$, and which are chosen with equal likelihood. All other players can each be of one type only, and hence we drop reference to their types.

- For a player $k \neq C$, the payoff ρ^k satisfies $\rho^k = r^k$, while

$$\rho^C(\uparrow, \cdot) = r^C(\cdot),\ \rho^C(\downarrow, \cdot) \equiv 0$$

  That is, all players other than C have the payoffs defined in Section 4; this is also true of C of type $\uparrow$, while C of type $\downarrow$ is an indifferent type.

- Define a positive sequence $(p_n)_{n=0}^{\infty}$ by:

$$p_{n+1} = \begin{cases} \frac{1}{2} & \text{if } n = 0 \\ \frac{p_n}{1-\delta_n^3} & \text{if } n > 0 \end{cases} \tag{8.2}$$

- Now, define the strategies $(\eta^j)_{j\in\mathcal{P}}$ in the following way (recall that we denote a mixed action in $\Delta(\{L, R\})$ by a single number in $p \in [0, 1]$, the probability of L, instead of $(p, 1-p)$):

  - $\eta^D = \tau^D$, i.e., D plays under η as he would in τ.
  - For $h \in \cup_t H_t$,

$$\eta^C(t^C, h) = \begin{cases} L & \text{if } \forall n, t \neq S_n \\ R & \text{if } t = S_n, t^C = \uparrow \\ \frac{\delta_n^3}{1-p_n} & \text{if } t = S_n, t^C = \downarrow \end{cases} \tag{8.3}$$

    Observe then that $\eta^C(\uparrow, h) = \tau^C(h)$, i.e., C of type $\uparrow$ plays as he would in τ.

  - For $h = (a_1, \ldots, a_{t-1}) \in \cup_t H_t$,

$$\eta^{A_j}(h) = \begin{cases} L & \text{if } \forall n, t \neq S_n \\ \frac{1}{2} + \delta_n & \text{if } t = S_n \text{ and } \forall k < n, a^C_{S_k} = R \\ \frac{1}{2} + \frac{\delta_n}{\sqrt[3]{1-p_n}} & \text{if } t = S_n \text{ and } \exists k < n, a^C_{S_k} = L \end{cases} \tag{8.4}$$

  Observe that as long as the players $A_1, \ldots, A_4$ observe C playing R at all stages in (S_n), they can continue to retain their uncertainty of the type of C, and play as they would in τ.

Recall the notations $G(\uparrow), G(\downarrow)$ which denote the stage game in which the type $\uparrow, \downarrow$, respectively, has been chosen and made public. Only C's payoff differs between the games. To show Result 2.6, we reason as follows: as long as C only plays R at the stages in (S_n), his type remains unknown but his *expected mixed action* will be precisely what is dictated by the beliefs of $\tau = (\tau^{j,k})_{j,k\in\mathcal{P}}$ defined in Section 5 (see Lemma 8.3). This will be the case forever when the type is $\uparrow$. Hence, in this case, the other players play an equilibrium response to C that is precisely what they would play under τ; and since C only needs to coordinate with D (his payoffs are not affected by $A_1, \ldots, A_4$), C is best-replying as well. If the type is $\downarrow$, C is indifferent anyway, and if he ever plays L when $\uparrow$ would not have, his type is revealed and the other players move to a revised equilibrium response in what then becomes a perfect information game (see Lemma 8.5). Hence, η is a Bayesian equilibrium. It also follows easily (Lemma 8.4) that, conditional on $\uparrow$, η and τ induced the same distribution on plays. We now formalise these; first, a technical issue:

Lemma 8.1. $\frac{1}{2} \leq p_n \leq \frac{2}{3}$ *for all* $n \in \mathbb{N}$.

Hence, since $\delta_n < \frac{1}{2\sqrt[3]{3}}$, and in particular $\frac{\delta_n^3}{1-p_n} \leq \frac{(\frac{1}{2\sqrt[3]{3}})^3}{\frac{1}{3}} = \frac{1}{2^3}$, the strategies given by (8.3) and (8.4) are well-defined.

Proof. This follows by (8.1), since one shows inductively that:

$$p_{n+1} = \frac{1}{2} \prod_{k \leq n} \frac{1}{1-\delta_k^3}$$

$\square$

Lemma 8.2. $p_n = P_\eta(t^C = \uparrow \mid R_n)$ *where*

$$R_n = \{(a_1, \dots, a_{S_n - 1}) \in H_{S_n} \mid \forall k < n, a^C_{S_k} = R\} \tag{8.5}$$

I.e., the probability that the other players associate with the type of C being $\uparrow$, given that only R has been played by him at each stage of the form S_k for $k < n$, is p_n. Note that $R_1 = \{\emptyset\}$, the empty history.

Proof. Observe that by the definition of η,

$$P_\eta(R_{n+1} \mid R_n \cap t^C = \downarrow) = 1 - \frac{\delta_n^3}{1-p_n},\ P_\eta(R_{n+1} \mid R_n \cap t^C = \uparrow) = 1$$

Assume inductively $p_n = P_\eta(t^C = \uparrow \mid R_n)$: This clearly holds for $n = 1$, since $P_\eta(t^C = \uparrow \mid \{\emptyset\}) = P_\eta(t^C = \uparrow) = \frac{1}{2} = p_1$. Using Bayes rule, since $R_{n+1} \subseteq R_n$,

$$\begin{aligned} &P_\eta(t^C = \uparrow \mid R_{n+1}) \\ &\quad = \frac{P_\eta(R_{n+1} \mid R_n \cap t^C = \uparrow) \cdot P_\eta(t^C = \uparrow \mid R_n)}{P_\eta(R_{n+1} \mid R_n \cap t^C = \uparrow) \cdot P_\eta(t^C = \uparrow \mid R_n) + P_\eta(R_{n+1} \mid R_n \cap t^C = \downarrow) \cdot P_\eta(t^C = \downarrow \mid R_n)} \\ &\quad = \frac{1 \cdot p_n}{1 \cdot p_n + (1 - \frac{\delta_n^3}{1-p_n}) \cdot (1-p_n)} = \frac{p_n}{1-\delta_n^3} = p_{n+1} \end{aligned}$$

as required. $\square$

Lemma 8.3. *For all* $n \in \mathbb{N}$, *all player* $k \neq C$, *and any* $h \in R_n$, $\eta^k(h)$ *is a best-response to* $\tilde{\tau}^k_{S_n}$ *(the beliefs of* k *at stage* S_n*) in the stage games* $G(\uparrow), G(\downarrow)$.

Proof. Indeed, by the previous lemma and the definition of $\tau^{k,C}_{S_n}$,

$$
\begin{aligned}
P_\eta(a^C_{S_n} = L \mid R_n) &= p_n \cdot \eta^C_{S_n}(\uparrow)[L] + (1-p_n)\eta^C_{S_n}(\downarrow)[L] \\
&= p_n \cdot 0 + (1-p_n)\frac{\delta_n^3}{1-p_n} = \delta_n^3 = \tau^{k,C}_{S_n}
\end{aligned}
$$

Since for any $h, h' \in R_n$, $\eta^C(t^C, h) = \eta^C(t^C, h')$ for either value of t^C, we have

$$\tau^{k,C}_{S_n} = P_\eta(a^C_{S_n} = L \mid h),\ \forall h \in R_n$$

To finish the proof, observe that for $h \in R_n$, $\tau^{-C}_{S_n} = \eta^{-C}(h)$ by comparing (8.4) with Section 5 and by $\eta^D \equiv \tau^D$, and recall $\tau^{-C}_{S_n}$ is an equilibrium response to the shared belief of the other players at stage S_n about C's action under τ. □

The following is clear from (8.3) and the definition of τ, since $P_\eta(h \mid\uparrow) > 0$ implies $h \in R_n$.

Lemma 8.4. *For each $t \in \mathbb{N}$ and each $h \in \cup_t H_t$ with $P_\eta(h \mid\uparrow) > 0$,*

$$P_\eta(a^k_t = L \mid h) = P_\tau(a^k_t = L \mid h) = 1$$

Lemma 8.5. *Denote the complement of R_n by*

$$V_n = \{(a_1, \ldots, a_{S_n - 1}) \in H_{S_n} \mid \exists k < n, a^C_{S_k} = L\}$$

i.e., the event that at some stage S_k for $k < n$, C plays L. Then

$$P_\eta(a^C_{S_n} = L \mid h) = \frac{\delta_n^3}{1-p_n},\ \forall h \in V_n \tag{8.6}$$

and if $h \in V_n$, then $P_\eta(t^C =\downarrow\mid h) = 1$ and $\eta(\downarrow, h)$ is an equilibrium of the stage game $G(\downarrow)$.

Proof. From (8.3), $P_\eta(\downarrow \mid h) = 1$ for $h \in V_n$, and then also from (8.3), (8.6) follows. Hence, just as when C is believed to play δ_n^3, it is an equilibrium response from $A_1, \ldots, A_n$ to play $\frac{1}{2} + \delta_n$, so when C is believed to play $\frac{\delta_n^3}{1-p_n}$, it is an equilibrium response from $A_1, \ldots, A_n$ to play $\frac{1}{2} + \frac{\delta_n}{\sqrt[3]{1-p_n}}$ (see proof of Lemma 5.1). □

Corollary 8.6. *$(\eta^k)_{k\in\mathcal{P}}$ is a Bayesian equilibrium.*

Proof. If $t^C =\uparrow$, then at each stage, C is best-replying in the stage game $G(\uparrow)$ to the actions of D, as C is playing a coordination game with D.[13] If $t^C =\downarrow$, C is indifferent anyway. For any $t \in \mathbb{N}$, the players other than C are always playing an equilibrium response to C's expected action in the stage game[14] $G(\uparrow)$ or $G(\downarrow)$ at stage t:

[13] D's actions are history-independent, playing R at stages $(S_n)_{n=1}^\infty$ and L otherwise.

[14] These games are equivalent for all players except C.

- If $\forall n, t \neq S_n$, this follows from Lemma 4.2.
- If $t = S_n$ and $h \in R_n$, this follows from Lemma 8.3.
- If $t = S_n$ and $h \in V_n$, this follows from Lemma 8.5

Although this reasoning concerns the *stage game*, it then follows that η gives an equilibrium response for the other players to C in the *repeated game*, since no player's strategy depends on any other player's previous actions other than C's and C's strategy depends on no player's previous actions (only on his type).[15]

□

The following corollary is immediate from Lemma 8.4:

Corollary 8.7. *P_τ is precisely the marginal of $P_\eta(\cdot\,|\uparrow)$ on H_∞.*

Recall that $G^\infty(\uparrow)$ is the infinitely repeated version of $G(\uparrow)$, in which the type $\uparrow$ is chosen and made common knowledge. Given Result 2.4:

Corollary 8.8. *For $P_\eta(\cdot\,|\uparrow)$-a.e. $h \in H_\infty$, for all $t \in \mathbb{N}$, and for all equilibrium σ of $G^\infty(\uparrow)$, we have $P_{\eta_{h|_t}}(\cdot\,|\uparrow) \perp P_\sigma$.*

Indeed, σ must also be an equilibrium of G^∞, for G defined in Section 4, since $G = G(\uparrow)$.

9 Contrast to Short-Run Merging

In [Kalai and Lehrer (1993), Sec. 7.1], an alternative notion of closeness, weaker than closeness in the total-variation norm, is discussed: Let σ, τ be two strategy profiles. For $\ell \in \mathbb{N}$, let $\pi^\ell : H_\infty \to H_{\ell+1}$ be the projection onto the first ℓ coordinates, and let $\pi^\ell_*(P_\tau)$, $\pi^\ell_*(P_\sigma)$ denote the measures $P_\tau \circ (\pi^\ell)^{-1}$, $P_\sigma \circ (\pi^\ell)^{-1}$, i.e., the marginals on $H_{\ell+1}$. The following is proven in [Kalai and Lehrer (1993), Sec. 7.1]:

Theorem 9.1. *Let $\tau = (\tau^{j,k})_{j,k \in \mathcal{P}}$ be such that everyone is best-replying to their beliefs. Denote by $\tilde{\tau}^j = (\tau^{j,k})_{k \in \mathcal{P}}$ the beliefs of Player j. Assume that for each $j \in \mathcal{P}$, $P_\tau \ll P_{\tilde{\tau}^j}$. Then for P_τ-a.e. $h \in H_\infty$, any $\varepsilon > 0$, and any $\ell \in \mathbb{N}$, there exists $T \in \mathbb{N}$ such that for all $t \geq T$, there is an (exact) equilibrium σ of G^∞ such that $||\pi^\ell_*(P_{\tau_{h|_t}}) - \pi^\ell_*(P_\sigma)|| < \varepsilon$.*

That is, it is possible after a long enough period of learning to find equilibria which are very similar to the true play *in the short run*; specifically, over any fixed-length horizon. However, the examples in our paper show that in general will *not* be possible to ensure that the predictions made will be similar to equilibria *in the long run.* Our papers shows a similar limitation in repeated

[15] In other words, each player other than C is facing a decision process which evolves independent of his actions; hence, if he is best-replying in the stage game at each stage, he is best-replying in the repeated game.

Bayesian games.

The difference can be illuminated by the following example, taken from [Jackson, Kalai, Smorodinsky (1999)]: A parameter θ is chosen uniformly in $[0,1]$ but we are not informed of its value; we then observe an infinite sequence of coin tosses, distributed[16] i.i.d. with parameter θ. After each toss we update our beliefs on the future. If we wait long enough, our conditional distribution over the true value of θ will become very 'narrow', which will allow us to make good predictions over short run horizons. However, in the strong convergence notion of total-variation – so seemingly appealing given the success of [Blackwell and Dubins (1962)] – this learning will always be insufficient for long-run predictions.

We give another example: A sequence of parameters $(\theta_n)_{n=1}^{\infty}$ is chosen independently, where θ_n takes values $\frac{1}{2} \pm \frac{1}{2\sqrt{n}}$ with equal probabilities. The sequence $(\theta_n)_{n=1}^{\infty}$ is not revealed to us. We view a sequence of coin tosses, conditionally independent on the choice of $(\theta_n)_{n=1}^{\infty}$, where the probability of heads on the n-th toss is θ_n.

It is easy to see in this case that, for any fixed $\ell \in \mathbb{N}$, we will eventually make very good predictions in the horizon of ℓ periods ahead; we don't even need to observe the coin tosses to make these predictions. However, any estimator $\hat{\theta}_n$ at stage n for the parameter used at stage n will, with probability $\frac{1}{2}$, be incorrect by an amount of at least $\frac{1}{2\sqrt{n}}$. In such a case, for example, any investor who tries to make successive bets using the Kelly criterion on a double-or-nothing bet with estimated parameter $\hat{\theta}_n$ at stage n will have his wealth grow much slower than an investor who knows the true underlying sequence $(\theta_n)_{n=1}^{\infty}$; this results from classical 'turnpike' theorems on optimal growth, e.g., [Breiman (1961), Thm. 3].

10 An Alternative Example

We present now an alternative example. The advantage of this example is that it is simpler in some respects, although shares a similar theme. A disadvantage of this example is that it cannot be modelled into the framework of Bayesian games, as we had done for the other construction in Section 8.[17] We do not present a complete proof that convergence to equilibria does not occur, rather we state parallels of Lemma 4.1 and Proposition 6.1, and the proof from there continues very similarly.

[16] The independence of the coin tosses is conditional on the choice of parameter θ.

[17] The reason is that, since types must be independent in the Bayesian games model in order to apply Theorem 2.5, it is always true that any two players must, at any given time, have the same belief about the type (and hence the actions) of any third player. In this example, this is not the case, and this disagreement is crucial for the success of the example.

There are seven player, $\mathcal{P} = \{A, B_1, B_2, C_1, C_2, C_3, D\}$. Each has actions $\{L, R\}$. The payoffs are given by:

- For each player $p \in \{C_1, C_2, C_3, A\}$, we have

$$r^p(R, \cdot) = \begin{cases} 0 & \text{if } a^D = R \text{ or } a^p = R \\ 1 & \text{if } a^D = L \text{ and } a^p = L \end{cases}$$

  When D plays L, these players prefer L; otherwise, they are indifferent. Denote $\Delta^0(x) = \prod_{j=1,2,3}(x^{C_j} - \frac{1}{2})$, $\Delta^A(x) = x^A - \frac{1}{2}$.

- For $j = 1, 2$,

$r^{B_j}(\cdot, x^{-B_j}) =$

	$x^D = L$	$x^D = R$
L	2	$\Delta^A(x) + (-1)^j \Delta^0(x)$
R	0	0

- $r^D(\cdot, x^{-D}) =$

L	$4x^A - 3$
R	0

  Note that A, D are playing a sort of coordinate game between them.

Again, as in Lemma 4.2, the profile $\overline{L}$ in which $\overline{L}^k = L$ for all $k \in \mathcal{P}$ is an equilibrium. The parallel of Lemma 4.1 is:

Lemma 10.1. *For each $0 < \delta < \frac{1}{4}$, there exists $\eta > 0$ such that there is no η-equilibrium z of G satisfying:*

- *For $j = 1, 2, 3$, $\frac{1}{2} + \frac{1}{2}\delta \leq z^{C_j}$.*
- *$z^{B_1}, z^{B_2} \in [\delta, 1 - \delta]$.*
- *$z^D = R$.*

Proof. As in the proof of Lemma 4.1, we would get an exact equilibrium z satisfying these conditions. Then $z^D = R$, and we have $\Delta^0(z) \neq 0$, so B_1, B_2 cannot be both indifferent between L, R. □

Similar to Section 5, let $(\delta_n)_{n=1}$, $(L_n)_{n=1}^{\infty}$ be a sequences satisfying (5.1) and (5.2), except with ε,ζ and the function T_0 being specified by Proposition 10.2 below instead of Proposition 6.1. Define $\tau_t = \overline{L} = (L, \ldots, L)$ if $\forall n \in \mathbb{N}, t \neq S_n$, while

$$\tau^p_{S_n} = \begin{cases} \frac{1}{2} + \delta_n & \text{if } p = C_1, C_2, C_3 \\ R & \text{if } p = D \\ \frac{1}{2} & \text{if } p = A, B_1, B_2 \end{cases}$$

The beliefs are defined as such:

- If $k \neq B_1, B_2$ or $m \neq A$, $\tau^{k,m} = \tau^m$ (i.e., all players except B_1, B_2 have correct beliefs; B_1, B_2 have correct beliefs about all others except A.)

- However, for $j = 1, 2$,

$$\tau_t^{B_j,A} = \begin{cases} \frac{1}{2} - (-1)^j \delta_n^3 & \text{if } t = S_n \\ L & \text{if } \forall n \in \mathbb{N}, t \neq S_n \end{cases}$$

We leave it to the reader to show the parallels of Lemmas 5.1 and 5.2 - that is, that each player is best-replying to his beliefs in each stage, and that the beliefs possess grain of truth. Now, like in Section 6, if we denote

$$H_{T+1}^{\alpha} = \{h = (a_1, \ldots, a_T) \in H_{T+1} \mid a_1^D = R \text{ and } \forall k \in \mathcal{P}, 2 \leq t \leq T, a_t^k = L\} \tag{10.1}$$

then:

Proposition 10.2. *There exists $\varepsilon > 0$ such that for each $0 < \delta$ small enough, there exists $T_0 = T_0(\delta) \in \mathbb{N}$ such that if $T \geq T_0$ then there does not exist an equilibrium profile σ which satisfies:*

- *For all $j = 1, 2, 3$, $\frac{1}{2} + \frac{1}{2}\delta \leq \sigma^{C_j}(\emptyset)$.*
- *$P_\sigma(H_{T+1}^{\alpha}) > 1 - \varepsilon$ and $P_\sigma(h) > \frac{1}{2}\zeta^{-1}$ for each $h \in H_{T+1}^{\alpha}$, where $\zeta = |H_{T+1}^{\alpha}| = 4^6$.*

Indeed, one proves parallels of Lemma 6.2, of Lemma 6.3 (for Player D), and of Lemma 6.4 (using η from Lemma 10.1 instead of Lemma 4.1); one then observes that if $\delta \leq \frac{1}{2}\zeta^{-1}$, then $P_\sigma(h) > \frac{1}{2}\zeta^{-1}$ for each $h \in H_{T+1}^{\alpha}$ implies that $\sigma^{B_1}(\emptyset), \sigma^{B_2}(\emptyset) \in [\delta, 1-\delta]$, and then derives a contradiction using Lemma 10.1 just as in the proof of Proposition 6.1.

Following the proof of this proposition, the proof that convergence to equilibria does not occur proceeds in the same manner as in Section 7.

11 Appendix A: Probabilistic Tools

Let[18] $X_1, X_2, \ldots$ be finite sets, $X = \prod_{n \in \mathbb{N}} X_n$, $\overline{X}_n = \prod_{k<n} X_k$. For a measure[19] $P \in \Delta(X)$, let P_n denote the marginal of P on $\overline{X}_{n+1}$. For $P, Q \in \Delta(X)$, we will say that P is *locally absolutely continuous w.r.t.* Q if $P_n \ll Q_n$ for all n. For each n, let $P_n[\cdot|\cdot]$ be the conditional on X_n w.r.t. $\overline{X}_n$: I.e., $P_n[\cdot \mid \overline{x}_n]$ is the distribution on X_n given $\overline{x}_n \in \overline{X}_n$. The distribution $P_n[\cdot \mid \overline{x}_n]$ is uniquely defined[20] if $P(\overline{x}_n) > 0$. A *version* of $P_n[\cdot \mid \cdot]$ is such a conditional distribution with $P[\cdot \mid \overline{x}_n]$ defined in this unique way when $P(\overline{x}_n) > 0$, and an arbitrary element of $\Delta(X_n)$ if $P(\overline{x}_n) = 0$.

[18]Some of the notation here is a repeat of Section 7.

[19]X is endowed with the Borel σ-algebra induced by the Tychonoff topology.

[20]For all $\overline{x}_n = (x_1, \ldots, x_{n-1}) \in X^*$, $P(\overline{x}_n) = \prod_{k<n} P[x_k \mid \overline{x}_k]$.

Proposition 11.1. *Let $(p_t)_{t=1}^{\infty}$, $(q_t)_{t=1}^{\infty}$ be sequences with $p_t, q_t \in \Delta(X_t)$, such that for all $t \in \mathbb{N}$, $p_t \ll q_t$. Let $P = \underset{t\in\mathbb{N}}{\otimes} p_t$, $Q = \underset{t\in\mathbb{N}}{\otimes} q_t$. Suppose $\sum_{t=1}^{\infty} ||p_t - q_t||_\infty < \infty$, and for some $\alpha > 0$, for all t, $p_t[x] > 0$ implies $q_t[x] \geq \alpha$. Then Q contains a grain of truth of P (i.e., $P \ll Q$ and $\frac{dP}{dQ}$ is bounded).*

Proof. For each element $\overline{x} \in X$ s.t. $P(\overline{x}_n) > 0$ for all n (which implies $Q(\overline{x}_n) > 0$ for all n), denote

$$T_n(\overline{x}) = \frac{P(\overline{x}_n)}{Q(\overline{x}_n)} = \prod_{k\leq n} \frac{p_k[x_k]}{q_k[x_k]}$$

Denote $Z = \lim_{n\to\infty} T_n(\overline{x})$; it is known that this limit exists Q-a.e., and if $Z < \infty$ Q-a.s., then $P \ll Q$ and $\frac{dP}{dQ} = Z$ (e.g., [Shiryaev (1995), Sec. VII.6, Thm. 1]). For Q-a.e. $\overline{x}$,

$$\begin{aligned}\ln(T_n(\overline{x})) &\leq \sum_{t=1}^{\infty} \sup_{x\in X_t, p_t[x]>0} \max(\ln(p_t[x]) - \ln(q_t[x]), 0) \\ &\leq \frac{1}{\alpha}\sum_{t=1}^{\infty} \sup_{x\in X_t, p_t[x]>0} \max(p_t[x] - q_t[x], 0) \leq \frac{1}{\alpha}\sum_{t=1}^{\infty} ||p_t - q_t||_\infty < \infty\end{aligned}$$

□

The following is known as the *Kakutani dichotomy* (the version for products of finite spaces); see [Shiryaev (1995), Sec. VII.6], Theorem 4, which generalises [Kakutani (1948)].

Theorem 11.2. *Let $P, Q \in \Delta(X)$ with P locally absolutely continuous w.r.t. Q. Then*

$$P \ll Q \iff P\Big[\sum_{n=1}^{\infty}\big(1 - \sum_{a\in X_n} \sqrt{P_n[a \mid \overline{x}_n] \cdot Q_n[a \mid \overline{x}_n]}\big) < \infty\Big] = 1$$

and

$$P \perp Q \iff P\Big[\sum_{n=1}^{\infty}\big(1 - \sum_{a\in X_n} \sqrt{P_n[a \mid \overline{x}_n] \cdot Q_n[a \mid \overline{x}_n]}\big) = \infty\Big] = 1$$

Note that the summand is well-defined - since if $Q(\overline{x}_n) = 0$ then $P(\overline{x}_n) = 0$ - and is seen to be non-negative.

Corollary 11.3. *Let $P, Q \in \Delta(X)$, and for each n, let $P_n[\cdot|\cdot], Q_n[\cdot \mid \cdot]$ be versions of the marginals and suppose that for P-a.e. $x \in X$,*

$$\sum_{n=1}^{\infty}\big(1 - \sum_{a\in X_n} \sqrt{P_n[a \mid \overline{x}_n] \cdot Q_n[a \mid \overline{x}_n]}\big) = \infty \tag{11.1}$$

Then $P \perp Q$.

Note that the power that Corollary 11.3 adds to the Kakutani criterion for the divergent case is that P need not be locally absolutely continuous w.r.t. Q.

Proof. Fix $\varepsilon > 0$, denote $\varepsilon_n = \frac{\varepsilon}{4^n \cdot |X_n|^2}$, and define for each n a new marginal by

$$Q_n^\varepsilon[\cdot \mid \cdot] = (1-\varepsilon_n)Q_n[\cdot \mid \cdot] + \varepsilon_n P_n[\cdot \mid \cdot]$$

Let $Q^\varepsilon \in \Delta(X)$ denote the induced measure. It's easy to show[21] that if (11.1) holds for some $x \in X$, then it also holds with $Q_n^\varepsilon[\cdot \mid \cdot]$ replacing $Q[\cdot \mid \cdot]$, and clearly P is locally absolutely continuous w.r.t. Q^ε. Applying the Kakutani criterion, we have $P \perp Q^\varepsilon$; repeating this for any $\varepsilon > 0$ gives a sequence $Q^\varepsilon \to Q$ in total-variation norm[22] and shows $P \perp Q$.[23] ☐

Theorem 11.4. *Let $P, Q \in \Delta(X)$ and suppose that for some versions of $P_n[\cdot \mid \cdot]$, $Q_n[\cdot \mid \cdot]$, it holds for those $x = (x_1, x_2, \ldots) \in X$ satisfying $P(x_1, \ldots, x_n) > 0$ for all n that*

$$\sum_{n=1}^{\infty} ||P_n[\cdot \mid x_1, \ldots, x_{n-1}] - Q_n[\cdot \mid x_1, \ldots, x_{n-1}]||^2 = \infty \tag{11.2}$$

Then $P \perp Q$.

Proof. Lemma 1 of [Lehrer and Smorodinsky (1997)] shows that if $p, q \in \Delta(A)$ for finite A, then $1 - \sum_{a \in A} \sqrt{p[a] \cdot q[a]} \geq \frac{||p-q||^2}{8}$. Hence the result follows from Corollary 11.3. ☐

We will also make use of the following:

Proposition 11.5. *Let $P, Q \in \Delta(X)$ and $T \in \mathbb{N}$. Denote $P_+ = \{\overline{x}_T \in \overline{X}_T | P(\overline{x}_T) > 0\}$ and similarly define Q_+. Suppose for each $\overline{x}_T \in P_+ \cap Q_+$, $P(\cdot \mid \overline{x}_T) \perp Q(\cdot \mid \overline{x}_T)$. Then $P \perp Q$.*

Proof. For each $\overline{x}_T \in P_+ \cap Q_+$, let $S_P(\overline{x}_T), S_Q(\overline{x}_T) \subseteq \prod_{n \geq T} X_n$ be disjoint such that $P(\overline{x}_T \times S_P(\overline{x}_T) \mid \overline{x}_T) = 1$, $Q(\overline{x}_T \times S_Q(\overline{x}_T) \mid \overline{x}_T) = 1$. For $\overline{x}_T \notin P_+ \cap Q_+$, take $S_P(\overline{x}_T) = S_Q(\overline{x}_T) = \prod_{n \geq T} X_n$. Then define,

$$S_P = \bigcup_{\overline{x}_T \in \overline{X}_T} (\overline{x}_T \times S_P(\overline{x}_T)), \quad S_Q = \bigcup_{\overline{x}_T \in \overline{X}_T} (\overline{x}_T \times S_Q(\overline{x}_T))$$

Then it is easy to check that

$$P(S_P) = Q(S_Q) = 1 - Q(S_P) = 1 - P(S_Q) = 1$$

☐

[21]E.g., using the inequality $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x-y|}$ for $x, y \geq 0$.

[22]It's easy to show that if $\mu, \nu \in \Delta(X)$,

$$||\mu - \nu|| \leq \sum_{n=1}^{\infty} \sum_{x_n \in X_n} \sup_{\overline{x}_n \in \overline{X}_n} |\mu_n[x_n \mid \overline{x}_n] - \nu_n[x_n \mid \overline{x}_n]|$$

[23]Indeed, this implies that for each $\varepsilon > 0$, there is $A \subseteq X$ with $P(A) < \varepsilon$, $Q(A) > 1 - \varepsilon$; this easily implies $P \perp Q$.

12 Appendix B: Proofs from Section 6

Suppose σ were an equilibrium satisfying the conditions given in Proposition 6.1, with $\varepsilon > 0$ and $T \geq T_0$ satisfying the inequalities there. Recall $\zeta = |H^{\alpha}_{T+1}| = 4^4$.

Observe that for $2 \leq t \leq T$, the elements of H^{α}_{t+1} are initial segments of the respective elements of H^{α}_{T+1}

For $t \in \mathbb{N}$ and $(a_1, a_2, \ldots) \in H_\infty$, denote

$$\overline{r}_t[a_1, a_2, \ldots] := \sum_{s \leq t} \beta^{s-1} r(a_s) \tag{12.1}$$

and

$$g(a_1, a_2, \ldots) = \sum_{t > T} \beta^{t-1} r(a_t)$$

Lemma 12.1. *For each $2 \leq t \leq T$ and each $h \in H^{\alpha}_t$,*

$$P_\sigma(H^{\alpha}_{T+1} \big| \{h\}) > 1 - 2\zeta\varepsilon$$

Proof. If not, let $2 \leq t \leq T$ and $h^* \in H^{\alpha}_t$ be such that $P_\sigma(H_{T+1} \backslash H^{\alpha}_{T+1} \big| \{h^*\}) \geq 2\zeta\varepsilon$. Since, by assumption, $P_\sigma(h^*) \geq \frac{1}{2}\zeta^{-1}$,

$$P_\sigma(H_{T+1} \backslash H^{\alpha}_{T+1}) \geq P_\sigma(H_{T+1} \backslash H^{\alpha}_{T+1} \big| \{h^*\}) \cdot P_\sigma(h^*) \geq (2\zeta\varepsilon) \cdot (\frac{1}{2}\zeta^{-1}) = \varepsilon$$

and therefore $P_\sigma(H^{\alpha}_{T+1}) \leq 1 - \varepsilon$, a contradiction. □

Proof. (Proof of Lemma 6.2) Suppose not; let $k \in \mathcal{P}$, $2 \leq t \leq T - T_1$ and $h^* \in H^{\alpha}_t$ such that $\sigma^k(h^*) < 1$. Let σ_L (resp. σ_R) be strategy profiles such that $\sigma^k_L(h^*) = L$ (resp. $\sigma^k_R(h^*) = R$) and agrees with σ otherwise.[24] We have:

$$E_{\sigma_R}[\overline{r}^k | \{h^*\}] = \overline{r}^k_{t-1}(h^*) + E_{\sigma_R}\big[\sum_{s=t}^{T} r^k(a_s)\beta^{s-1} + g^k(h) \mid \{h^*\}\big]$$

$$\leq \overline{r}^k_{t-1}(h^*) + 1 \cdot \beta^{t-1} + 2 \cdot \sum_{s=t+1}^{T} \beta^{s-1} + ||g||_\infty$$

where we have used the fact that $r^k(R, x^{-k}) \leq 1$ for any profile x, and $r^k \leq 2$. On the other hand, since for any $(a_1, \ldots, a_T) \in H^{\alpha}_{T+1}$, $r^k(a_t) = 2$ for all $2 \leq t \leq T$, and $r^k \geq -||r||_\infty$,

$$E_{\sigma_L}[\overline{r}^k | \{h^*\}] = \overline{r}^k_{t-1}(h^*) + E_{\sigma_L}\big[\sum_{s=t}^{T} r^k(a_s)\beta^{s-1} + g^k(h) \mid \{h^*\}\big]$$

$$\geq \overline{r}^k_{t-1}(h^*) + (2 \cdot P_{\sigma_L}\big(H^{\alpha}_{T+1} \big| \{h^*\}) - ||r||_\infty \cdot (1 - P_{\sigma_L}(H^{\alpha}_{T+1} \big| \{h^*\}))\big) \sum_{s=t}^{T} \beta^{s-1} - ||g||_\infty$$

$$\geq \overline{r}^k_{t-1}(h^*) + (2(1 - 2\zeta\varepsilon) - 2\zeta\varepsilon||r||_\infty) \sum_{s=t}^{T} \beta^{s-1} - ||g||_\infty$$

[24] That is, $\sigma^m_L = \sigma^m_R = \sigma^m$ for $m \neq k$, and $\sigma^k_L(q) = \sigma^k_R(q) = \sigma^k(q)$ for any $q \in \cup_{t \leq T} H_t$ except for the one case $q = h^*$.

where we have used Lemma 12.1, since $P_{\sigma_L}(H^\alpha_{T+1}\big|h^*) \geq P_\sigma(H^\alpha_{T+1}\big|h^*)$. In order to have $E_{\sigma_R}[\overline{r}^k] < E_{\sigma_L}[\overline{r}^k]$ (which implies $E_\sigma[\overline{r}^k] < E_{\sigma_L^k}[\overline{r}]$ and gives the desired contradiction to $\sigma^k(h^*) < 1$ since σ is an equilibrium), it suffices to have,

$$1 \cdot \beta^{t-1} + \sum_{s=t+1}^{T} 2\beta^{s-1} + ||g||_\infty < (2 - 2\zeta\varepsilon(2 + ||r||_\infty)) \sum_{s=t}^{T} \beta^{s-1} - ||g||_\infty$$

or equivalently,

$$2||g||_\infty < \beta^{t-1} - 2\zeta\varepsilon(2 + ||r||_\infty) \sum_{s=t}^{T} \beta^{s-1}$$

Since $||g||_\infty \leq \beta^T \frac{||r||_\infty}{1-\beta}$, $\sum_{s=t}^{T} \beta^{s-1} \leq \frac{\beta^{t-1}}{1-\beta}$, and $t \leq T - T_1$, it suffices to have

$$\beta^{T_1+1} \frac{||r||_\infty}{1-\beta} + \zeta\varepsilon(2 + ||r||_\infty) \frac{1}{1-\beta} < \frac{1}{2}$$

For this to hold, it suffices to require each term to be $< \frac{1}{4}$, which follow from (6.3) and (6.2). □

Proof. (Proof of Lemma 6.3) We deal with Player D; Player C follows similarly. Like in the previous proof, let σ_L (resp. σ_R) be the strategy profile such that $\sigma^D_L(\emptyset) = L$ (resp. $\sigma^D_R(\emptyset) = R$) and agrees with σ otherwise. Observe that $P_\sigma(H^\alpha_{T+1}) > 1 - \varepsilon$ implies $\sigma^C(\emptyset)[R] > 1 - \varepsilon$; and $P_{\sigma_L}(H^\alpha_{T+1}) = 0$ by definition. Hence

$$\begin{aligned} E_{\sigma_L}[r^D(a_1)] &= E_{\sigma_L}[r^D(a_1)|H^\alpha_{T+1}]P_{\sigma_L}(H^\alpha_{T+1}) + E_{\sigma_L}[r^D(a_1)|H_T \backslash H^\alpha_{T+1}](1 - P_{\sigma_L}(H^\alpha_{T+1})) \\ &\leq E_{\sigma_L}[r^D(a_1)|H^\alpha_{T+1}] \cdot 0 + (2 \cdot \sigma^C(\emptyset)[L] + 0 \cdot \sigma^C(\emptyset)[R]) \cdot (1 - 0) \leq 2\varepsilon < \frac{1}{4} \end{aligned}$$

by (6.2), which implies $\varepsilon < \frac{1}{8}$, and since $r^D \leq 2$. Hence, also since $r^D \leq 2$, we have

$$E_{\sigma_L}[\overline{r}^D] = E_{\sigma_L}[\sum_{s=1}^{T} r^D(a_s)\beta^{s-1} + g(h)] \leq \frac{1}{4} + 2\sum_{s=2}^{T} \beta^{s-1} + \beta^T \frac{||r||_\infty}{1-\beta}$$

Similarly, since $P_{\sigma_R}(H^\alpha_{T+1}) \geq P_\sigma(H^\alpha_{T+1}) > 1 - \varepsilon$ and $r^D \geq 0$, and $r^D(a_1) = 1$ if $(a_1, \ldots, a_T) \in H^\alpha_{T+1}$,

$$E_{\sigma_R}[r^D(a_1)] \geq E_{\sigma_R}[r^D(a_1)|H^\alpha_{T+1}]P_{\sigma_R}(H^\alpha_{T+1}) \geq 1 \cdot (1 - \varepsilon) \geq \frac{3}{4}$$

Therefore, since $r^D \geq 0$ (and hence $g^D \geq 0$), and for any $(a_1, \ldots, a_T) \in H^\alpha_{T+1}$, $r^k(a_t) = 2$ for all $2 \leq t \leq T$,

$$\begin{aligned} E_{\sigma_R}[\overline{r}^D] &= E_{\sigma_R}[\sum_{s=1}^{T} r^D(a_s)\beta^{s-1} + g^D(h)] \\ &\geq \frac{3}{4} + 2 \cdot P_{\sigma_R}(H^\alpha_{T+1}) \sum_{s=2}^{T} \beta^{s-1} \geq \frac{3}{4} + 2(1 - \varepsilon) \sum_{s=2}^{T} \beta^{s-1} \end{aligned}$$

where we have used again the fact that $P_{\sigma_R}(H^\alpha_{T+1}) \geq P_\sigma(H^\alpha_{T+1}) > 1-\varepsilon$. In order to have $E_{\sigma_R}[\overline{r}^D] > E_{\sigma_L}[\overline{r}^D]$ it is sufficient to require

$$\frac{3}{4} - 2\varepsilon \sum_{s=2}^{T} \beta^{s-1} > \frac{1}{4} + \beta^T \frac{||r||_\infty}{1-\beta}$$

which holds if

$$2\varepsilon \sum_{s=2}^{T} \beta^{s-1} + \beta^T \frac{||r||_\infty}{1-\beta} < \frac{1}{2}$$

and since $T \geq T_1$, this holds if

$$\varepsilon \frac{1}{1-\beta} + \frac{1}{2}\beta^{T_1} \frac{||r||_\infty}{1-\beta} < \frac{1}{4}$$

This will hold if both terms are less than $\frac{1}{8}$, which follows from (6.3) and (6.2). □

Proof. (Proof of Lemma 6.4) For each $u \in \{L,R\}^{\overline{\mathcal{P}}}$, where $\overline{\mathcal{P}} = \{A_1, A_2, A_3, A_4\}$, let σ_u be the strategy profile in which $\sigma_u^{\overline{\mathcal{P}}}(\emptyset) = u$, and σ_u agrees with σ otherwise.[25] $z = \sigma^{\overline{\mathcal{P}}}(\emptyset)$ must then be an equilibrium of the game with payoff

$$E_{\sigma_u}[\overline{r}^{\overline{\mathcal{P}}}] = r^{\overline{\mathcal{P}}}[u] + \xi(u)$$

where

$$\xi(u) = E_{\sigma_u}[\sum_{s=2}^{\infty} r^{\overline{\mathcal{P}}}(a_s)\beta^{s-1}]$$

If we can show that $||\xi(u) - \xi(v)||_\infty < \eta$ for any two $u, v \in \{L,R\}^{\overline{\mathcal{P}}}$, then $x = \sigma(\emptyset)$ will be an η-equilibrium in the stage game $r(\cdot)$, since $\sigma^{C,D}(\emptyset) = (R,R)$ by Lemma 6.3.

We have by Lemmas 6.2 and 6.3,

$$P_\sigma(\forall 2 \leq s \leq T - T_1, a_s = \overline{L}) = 1$$

and hence for $2 \leq s \leq T - T_1$,

$$P_\sigma(r^{\overline{\mathcal{P}}}(a_s) \equiv 2) = 1$$

Furthermore, since $P_\sigma(h) > 0$ for all $h \in H^\alpha_{T+1}$, $\sigma(\emptyset)[u] > 0$ and $\sigma(\emptyset)[v] > 0$ for any $u, v \in \{L,R\}^{\overline{\mathcal{P}}}$, hence $P_{\sigma_u} \ll P_\sigma$ and $P_{\sigma_v} \ll P_\sigma$. Hence, for $2 \leq s \leq T - T_1$,

$$E_{\sigma_u}[r^{\overline{\mathcal{P}}}(a_s)] = E_{\sigma_v}[r^{\overline{\mathcal{P}}}(a_s)] \equiv 2$$

Hence, to show that $||\xi(u) - \xi(v)||_\infty < \eta$, it's enough to require that

$$(\frac{||r||_\infty \cdot \beta^{T-T_1}}{1-\beta} =) \sum_{s=T-T_1+1}^{\infty} ||r||_\infty \beta^{s-1} < \frac{\eta}{2}$$

which follows from (6.3), since $T - T_1 \geq T_1$. □

[25] That is, $\sigma_u^{C,D}(\emptyset) = \sigma^{C,D}(\emptyset)$ and if $q \in \cup_{t=2}^{\infty} H_t$, $\sigma_u(q) = \sigma(q)$.

References

[Blackwell and Dubins (1962)] Merging of Opinions with Increasing Information, *The Annals of Mathematical Statistics*, Vol. 33, 882–886.

[Breiman (1961)] Optimal Gambling Systems for Favorable Games, *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob.*, Vol. 1, 65–78.

[Foster and Young (2001)] On the Impossibility of Predicting the Behavior of Rational Agents, *Proc. Nat. Acad. Sci.*, 98, 12848–12853.

[Gilli (2001)] A General Approach to Rational Learning in Games, *Bull. Econ. Res.*, 53, 275–303.

[Jackson, Kalai, Smorodinsky (1999)] Bayesian representation of stochastic processes under learning: de Finetti revisited. *Econometrica*, 67, 875-893.

[Jeitschko (1998)] Learning in Sequential Auctions, *Southern Econ. J.*, 65, 98–112.

[Jordan (1991)] Bayesian Learning in Normal Form Games, *Games Econ. Behav.*, 3, 60–81.

[Jordan (1993)] Three Problems in Learning Mixed–Strategy Nash Equilibria, *Games. Econ. Behav.*, 5, 368–386.

[Jordan (1995)] Bayesian Learning in Repeated Games, Games Econ. Behav., 9, 8–20.

[Kakutani (1948)] On Equivalence of Infinite Product Measures, *Annals of Mathematics*, 49, 214–224.

[Kalai and Lehrer (1993)] Rational Learning Leads to Nash Equilibrium, *Econometrica*, 61, 1021–1045.

[Kalai and Lehrer (1995)] Subjective Games and Equilibria, *Games. Econ. Behav.*, 8, 123–63.

[Lehrer and Smorodinsky (1996)] Compatible Measures and Merging, *Math. Op. Res.*, 21, 697–706.

[Lehrer and Smorodinsky (1997)] Repeated Large Games with Incomplete Information, *Games. Econ. Behav.*, 18, 116–134.

[Miller and Sanchirico (1997)] Almost Everybody Disagrees Almost All the Time: The Genericity of Weakly Merging Nowhere, Columbia University Discussion Paper No. 9697–25.

[Miller and Sanchirico (1999)] The Role of Absolute Continuity in Merging of Opinions and Rational Learning, *Games Econ. Behav.* 29, 170–190.

[Nachbar (1997)] Prediction, Optimization, and Learning in Repeated Games, *Econometrica*, 65, 275–309.

[Norman (2012)] Almost-Rational Learning of Nash Equilibrium without Absolutely Continuity, *Economics Series Working Papers*, 602, Univ. Oxford, Dep. Econ.

[Nyarko (1998)] Bayesian Learning and Convergence to Nash Equilibria without Common Priors, *Econ. Theory*, 11, 643–55.

[Sandroni (1998)] Necessary and Sufficient Conditions for Convergence to Nash Equilibrium: The Almost Absolute Continuity Hypothesis, *Games Econ. Behav.*, 22, 121–147.

[Sandroni and Smorodinsky (1999)] The Speed of Rational Learning, *Int. J. Game Theory*, 28, 199–210.

[Shiryaev (1995)] *Probability*, 2nd Ed., Springer.