The Possibility of Bayesian Learning in Repeated Games[*]

Thomas W. L. Norman

Magdalen College, Oxford

September 16, 2022

Abstract

In infinitely repeated games, Nachbar (1997, 2005) has shown that Bayesian learning of a restricted strategy set is inconsistent; the beliefs required to learn any element of such a set will lead best responses to lie outside of it in most games. But I establish here that Nash convergence of Bayesian learning requires only that optimal play (rather than any possible play) is learnable, and an appropriately modified notion of learnability is consistent in many of the games to which Nachbar's result applies. This means that rational learning of equilibrium is possible in an important class including coordination games, which I illustrate with two examples of positive learning results. *Journal of Economic Literature* Classification: C72; C73; D83.

Key Words: Repeated games; Nash equilibrium; Bayesian learning; rational learning; consistency.

[*] Email: thomas.norman@magd.ox.ac.uk.

1 Introduction

There are many possible ways in which players might learn to play an infinitely repeated game. But whichever way they choose, if they satisfy the Savage (1954) axioms, their play will respond optimally to subjective beliefs about opponents' strategies, and those beliefs will be updated in light of observed play in accordance with Bayes' law. Kalai and Lehrer (1993a) establish that such players will eventually approximate a Nash equilibrium of the repeated game, provided that this equilibrium play is absolutely continuous with respect to their beliefs. However, this "grain of truth" condition—whereby players must put positive probability on the eventual play from the outset—has been shown to be quite demanding; it is not, for instance, satisfied under pure strategies and atomless prior beliefs, even if the beliefs have "full support" on the repeated-game strategies.[1] Positive results are available if one adds further structure, such as (incomplete) payoff information with common priors (Jordan 1995) or with mutually absolutely continuous priors (Nyarko 1998), but without a "grain of truth" purely Savage–Bayesian learners are liable to spend eternity in disequilibrium.[2] And essentially, the "grain of truth" is a weakened form of the coordination in beliefs required by Nash equilibrium.

Nevertheless, perhaps beliefs might be uncoordinated but limited to a reasonable subset of the repeated-game strategy space, within which play might be "learnable". Nachbar (1997, 2005) argues that this works essentially only in games that are strategically simple in the sense of having a stage-game weakly dominant action; for other stage games, if the restricted strategy space is "broad enough" to avoid implicit coordination, a prior belief guaranteeing the learnability of any continuation play will generally lead ε-best responses to lie outside of the restricted strategy space for some such play. This "inconsistency" result suggests that the scope of Bayesian learning of Nash equilibrium is quite limited; more limited, for instance, than that of various sub-rational learning processes from evolutionary game theory (see, e.g., Sandholm 2011), where Nash equilibria of coordination games are eminently learnable.

In this paper, I show that this negative conclusion can be moderated, because Nash convergence of Bayesian learning does not require Nachbar's notion of learnability; rational learning requires only that *optimal* play be learnable, rather than *any* possible continuation

[1] Relatedly, Miller and Sanchirico (1997, 1999) critique the genericity and implicit coordination of absolute continuity, and Foster and Young (2001) exhibit the impossibility of learning certain mixed equilibria of standard Bayesian games. Kalai and Lehrer's (1993a) "grain of truth" condition is slightly stronger than absolute continuity.

[2] A framework for the analysis of equilibrium under 'misspecified' models is offered by Esponda and Pouzo (2016).

play. I offer a modified notion of "optimizing learnability" that captures this idea and is sufficient for approximate Nash convergence (Theorem 1 below).[3] I then show that such optimizing learnability may be reconciled with "broad enough" strategy sets in a wider class of games than Nachbar's consistent class (Propositions 1 and 2), including many repeated coordination games for instance. I offer two examples of positive learning results by way of illustration; one of 'dumb' learning under a uniform prior and bounded memory (anticipated in Noguchi 2015b), and the other of 'smart' learning under Sandroni's (2000) belief in "strict reciprocity".

Nachbar (2005, pp. 459–60) offers a simple example to illustrate his inconsistency: each player has two stage-game actions and believes that his opponent's strategy is i.i.d. for sure, but then an i.i.d. response is not even approximately optimal (under a belief that guarantees that any i.i.d. play is learnable) unless players have weakly dominant stage-game actions. Hence, Bayesian learning "within" the set of i.i.d. strategies is not possible. However, notice that i.i.d. play will eventually be optimal if players are coordinating in repeated coordination games, by contrast with repeated matching pennies. The idea in this paper is that, in many games without weakly dominant stage-game actions, there are nonetheless beliefs that generate learnable optimal play that is consistent with those beliefs; such beliefs may not be able to learn *any* path of play, but they can learn the path of *rational* play that they generate.

I now proceed informally to motivate this idea, before going on formally to: construct the model in Section 2; define the key concept of "optimizing learnability" in Section 3; show that it is inconsistent with "broad enough" strategy sets in a narrower class of games than Nachbar's learnability (Proposition 1); and show in Section 4 the sufficiency of optimizing learnability for Bayesian learning in repeated games both in general (Theorem 1) and for two particular coordination-game examples.

Informal Motivation The idea of Nachbar's (1997, 2005) inconsistency result is roughly as follows. A Nash equilibrium (with degenerate beliefs on the true strategies) is trivially "learnable", in the sense that the players (immediately) learn the continuation play. It is also trivially "consistent", in the sense that each player's strategy is a best response to his belief in any possible continuation. But it also clearly presupposes a great deal of coordination amongst the players in the veracity of their beliefs, to which criticism the

[3] Convergence to the set of exact Nash equilibria cannot be guaranteed under Kalai and Lehrer's strong notion of convergence (Levy 2015), but it is possible under weaker notions (Kalai and Lehrer 1993a §7.1, Sandroni 1998). Although the notion of convergence that I use is weaker than that of Kalai and Lehrer (1993a), it is sufficient for the discounting case (Kalai and Lehrer 1994). I employ a stronger notion of learning a particular path of play than Nachbar (2005), but require it to hold along fewer play paths.

	H	T
H	0 0	2 2
T	1 1	0 0

Figure 1: game in class MM

literature on learning in games is in large part a response. Nachbar considers whether Bayesian learning can remove this coordination by providing beliefs that are guaranteed to learn the continuation play of any of a broad set of strategy vectors, within which each player has a strategy that is a best response to his belief in any possible continuation. His inconsistency result shows that learning cannot accomplish this aim for the broad class (MM) of stage games with a minmax payoff in excess of their pure maxmin payoff; any set of strategy vectors that is broad enough to remove implicit coordination (in a precise sense) must fail to be either learnable or consistent in such games. In the class MM then, it is not possible for players' beliefs to be confined to a set of strategy vectors that is: (a) sufficiently small to allow them to learn continuation play; and (b) sufficiently large both to remove implicit coordination and to contain optimal responses for the players.

Consider the intuition for this result in Nachbar's i.i.d. strategies example above. Given the players' certainty that opponents' strategies are i.i.d., it can be shown that beliefs with full support on the i.i.d. strategy space guarantee the learnability of continuation play under any of its elements, notwithstanding their lack of a grain of truth. But for any given strategy f_i of a player i in games in class MM, there exists an opponent's strategy generating continuation play from which i would like to deviate (what Nachbar calls an "evil twin"). For instance, if player i plays H in every repetition of the game in Figure 1, then his opponent playing likewise will lead to infinite repetition of (H, H). Since a full-support prior belief guarantees that this continuation play is learned, player i cannot be optimizing in this continuation. And since such an evil twin can be found for any i.i.d.

strategy in the class MM of games, it follows in this class that the set of such strategies does not contain a strategy that is a best response for player i in every possible continuation—i.e. it is not consistent.

At a fundamental level then, it is difficult to guarantee both learning and optimization in all possible continuations. But should we be concerned with *all* possible continuations in this way? After all, infinite repetition of (H, H) in the game of Figure 1 calls for rather odd behavior on the part of the evil twin; he is evil not only to his twin, but also to himself! If both twins were optimizing with respect to their beliefs, and if they also both learned the continuation path of play, then this state of affairs would have zero probability. Hence it seems unnecessarily demanding to require that such a path be learnable; Bayesian learning of Nash equilibrium can certainly proceed in the absence of this requirement, as long as paths generated by *optimizing* play are learnable. Nachbar's evil twin is insufficient to violate consistency under this weaker form of learnability; to violate consistency, the evil twin would have to be optimizing—an "evil genius", if you will—as for instance with infinite repetition of the profile (H, T) in the game of matching pennies.

2 The Model

Now let us sketch the formal details of the learning model. There is a finite, two-player stage game with finite action sets Σ_1, Σ_2, and payoff functions $u_i : \Sigma \to \mathbb{R}$, $i = 1, 2$, $\Sigma \equiv \Sigma_1 \times \Sigma_2$. Without loss of generality, we may normalize the payoffs to be contained in the unit interval. This stage game is infinitely repeated in discrete time $t \in \mathbb{N}$, with each player knowing his own payoff function and public observation of play. As in Nachbar (2005), the results could be extended to games with more than two players, at the cost of some expositional simplicity.

Let $H_t \equiv \Sigma^t$ be the set of *histories* of actions taken in periods 1 through t, with $H \equiv \bigcup_{t \in \mathbb{N}} H_t$. The *realized play path* is the infinite vector $z \in \Sigma^\infty$ of the players' actions. For every history $h \in H_t$, a *cylinder with base on* h is the set $C(h) = \{z \in \Sigma^\infty \mid z = (h, \ldots)\}$ of all realized play paths whose t initial elements $z(t)$ coincide with h. Let $\mathscr{F}_t$ be the σ-algebra on Σ^∞ whose elements are all finite unions of cylinders with base on elements of H_t. We then have a filtration

$$\mathscr{F}_0 \subset \cdots \subset \mathscr{F}_t \subset \cdots \subset \mathscr{F},$$

where $\mathscr{F}_0$ is the trivial σ-algebra and $\mathscr{F}$ is the σ-algebra generated by the algebra of histories $\mathscr{F}^0 \equiv \bigcup_{t \in \mathbb{N}} \mathscr{F}_t$. The inclusion of $\mathscr{F}_t$ in $\mathscr{F}_{t+1}$ arises because any given event in period t is a union of events in period $t+1$.

A (behavior) *strategy* $f_i : H \to \Delta(\Sigma_i)$ for player i is a function from the set of all possible histories into the simplex of his action mixtures $\Delta(\Sigma_i)$; the space of all such strategies is $F_i \equiv \Delta(\Sigma_i)^\infty$, endowed with the product Borel σ-algebra, and the space of strategy vectors is $F \equiv F_1 \times F_2$. With a minor abuse of notation, the set G_i of *pure strategies* $g_i : H \to \Sigma_i$ may be considered a subset of F_i. Note that any f_i induces a strategy, f_i^l, in the corresponding l-fold repeated game. The f_i^l is simply the restriction of f to the smaller domain, H_l, and it is called the *l-truncation* of f. Given a strategy f_i, $t \in \mathbb{N}$ and $h \in H_t$, the *induced strategy* $f_i|_h$ is defined by

$$f_i|_h(h') = f_i(hh'), \quad \text{for any } h' \in H_r,\ r \in \mathbb{N},$$

where hh' is the concatenation of h with h', i.e. the history of length $t+r$ whose first t elements coincide with h, followed by the r elements of h'. Let $f_i|_h(a_i)$ denote the probability that $f_i|_h$ prescribes for the action $a_i \in \Sigma_i$.

The vector $f = (f_1, f_2)$ of behavior strategies induces a probability measure μ_f on the cylinder sets, with $\mu_f(C(h))$ giving the probability of the history h; $\mu_f(C(\emptyset)) = 1$ and $\mu_f(C(ha)) = \mu_f(h) \times_i f_i|_h(a_i)$, $h \in H$. The Kolmogorov Extension Theorem then delivers a unique extension of the probability measure μ_f from the $\mathscr{F}_t$'s to $\mathscr{F}$. Each player i also has a *belief*, which is a probability measure π^i over the opponent's behavior strategy. By Aumann's (1964) infinite-extensive-form version of Kuhn's (1953) theorem, this belief has an equivalent single behavior-strategy vector $f_{\pi^i} \equiv f^i = (f_1^i, f_2^i)$ (see Kalai and Lehrer 1993a, p. 1031), with $f_i^i = f_i$ (since a player is assumed to know his own strategy); let $-i$ be player i's opponent, and f_{-i}^i be the restriction of f^i to i's opponent's strategy space.

Letting $\lambda_i \in (0,1)$ be individual i's discount factor, player i's expected utility over all possible play paths is

$$U_i(f_i, f_{-i}^i) \equiv \int_{\Sigma^\infty} (1-\lambda_i) \sum_{t=1}^{\infty} \lambda_i^t u_i(z(t)) d\mu_{f^i},$$

and f_i is a *best response to* f_{-i}^i if

$$U_i(f_i, f_{-i}^i) \geq U_i(f_i', f_{-i}^i), \quad \forall f_i' \in F_i.$$

Given a small $\varepsilon > 0$, f_i is an *ε-best response to* f_{-i}^i if

$$U_i(f_i, f_{-i}^i) \geq U_i(f_i', f_{-i}^i) - \varepsilon, \quad \forall f_i' \in F_i.$$

A strategy vector f is an ε-*Nash equilibrium* if each f_i is an ε-best response to f_{-i}.

In order to analyze convergence of play and beliefs, we need a notion of the distance between probability measures on Σ^∞. I will often employ Sandroni's (1998) notion of weak closeness of strategy vectors: a probability measure μ is *weakly* ε*-close* to $\tilde{\mu}$ if

$$d(\mu, \tilde{\mu}) \equiv \sum_{k=1}^{\infty} 2^{-k} \left(\sup_{A \in \mathscr{F}_k} |\mu(A) - \tilde{\mu}(A)| \right) \leq \varepsilon,$$

where $\mathscr{F}_k$ is the σ-algebra of k-length histories defined earlier. Given two strategy vectors f and g (belonging to the general strategy-vector space F), f *plays weakly* ε*-like* g if μ_f is weakly ε-close to μ_g. The product topology on F is metrized by d. Intuitively, if two strategy vectors play weakly ε-like one another, then they induce two probability measures on histories that assign similar probabilities for all measurable events except perhaps those that may only be observed in the distant future (i.e. only in later $\mathscr{F}_k$'s).

If beliefs converge to strategies in the metric d, we have the case of "weak merging" of beliefs with true play. Nachbar's (2005) inconsistency result, however, employs a weaker notion of merging than this. Say that $\mathbb{N}^\circ \subseteq \mathbb{N}$ has *density* 1 if and only if

$$\lim_{n \to \infty} \frac{|\{1, \ldots, n\} \cap \mathbb{N}^\circ|}{n} = 1.$$

Definition 1 *Given beliefs π^1, π^2 and the belief-equivalent strategies f_2^1, f_1^2, player i* weakly learns to predict the path of play *generated by the behavior strategy vector f if and only if the following conditions hold:*

i. for any history $h \in H$, $\mu_f(C(h)) > 0$ implies $\mu_{(f_i, f_{-i}^i)}(C(h)) > 0$; and

ii. for any $\eta > 0$ and μ_f-almost any path of play z, there is a set $\mathbb{N}^P(\eta, z) \subseteq \mathbb{N}$ of density 1 such that, for any $n \in \mathbb{N}^P(\eta, z)$ and any $a_{-i} \in \Sigma_{-i}$,

$$\left| f_{-i}|_{z(n)}(a_{-i}) - f_{-i}^i|_{z(n)}(a_{-i}) \right| < \eta.$$

Given beliefs π^1 and π^2: a set $\widehat{F} \subseteq F$ of strategy vectors satisfies weak learnability *if and only if, for any $f_1 \in \widehat{F}_1$ and any $f_2 \in \widehat{F}_2$, each player weakly learns to predict the path of play; $\widehat{F}$ satisfies* pure weak learnability *if and only if the set $\widehat{G}$ of pure strategies in $\widehat{F}$ is weakly learnable.*

For given beliefs then, any element of a weakly learnable set of strategy vectors is close to beliefs for all but finitely many periods. This corresponds in the repeated-game setting to

Lehrer and Smorodinsky's (1996) more general concept of "almost weak merging", which is implied by "weak merging" in the metric d above. If requirement *ii* in Definition 1 is replaced by the stronger condition that, for any $\eta > 0$ and μ_f-almost any path of play z, there exists an $n(z) \in \mathbb{N}$ such that $f|_{z(n)}$ plays weakly η-like $f^i|_{z(n)}$ for all $n \geq n(z)$, then I will say that player i *learns to predict the path of play.*

In order to restrict the strategy set without implicit coordination on particular outcomes, we require some properties of such a restricted set, and here I follow Nachbar (2005), who provides further discussion.

Definition 2 (Nachbar 2005) *A set $\widehat{F} \subseteq F$ of strategy vectors satisfies* caution and symmetry (CS) *if and only if, for any pure strategy g_1 belonging to the set $\widehat{G}_1$ of pure strategies in $\widehat{F}_1$ and for any function $\gamma_{12} : \Sigma_1 \to \Sigma_2$ there is a pure strategy g_2 belonging to the set $\widehat{G}_2$ of pure strategies in $\widehat{F}_2$ such that the following is true. Let z be the path of play generated by (g_1, g_2). There is a set $\mathbb{N}^\gamma(z) \subseteq \mathbb{N}$ of density 1 such that for any $n \in \mathbb{N}^\gamma(z)$,*

$$g_2(z(n)) = \gamma_{12}(g_1(z(n))).$$

An analogous statement holds for $g_2 \in \widehat{G}_2$ and $\gamma_{21} : \Sigma_2 \to \Sigma_1$.

Definition 3 (Nachbar 2005) *A set $\widehat{F} \subseteq F$ of strategy vectors satisfies* pure strategies (P) *if and only if there is an $\nu > 0$ such that, for each i, the following is true. Consider any $f_i \in \widehat{F}_i$. There is a pure strategy $g_i \in \widehat{F}_i$ such that, for any history h, if $g_i(h) = a_i$, then $f_i|_h(a_i) > \nu$.*

A strategy set satisfying both conditions CS and P is said to satisfy CSP.

3 A Weakened Inconsistency Result

In this section, I alter Nachbar's (2005) notion of learnability, and show that the modified "optimizing learnability" limits the games and discount factors to which his inconsistency result applies. Thus, whilst impossibility problems remain for Bayesian learning, they present themselves under low discount factors and in a narrower class of games including repeated matching pennies but excluding many coordination games. Persistent inconsistency problems are associated with fully mixed equilibria, which accords with the impossibility results of Foster and Young (2001).

Nachbar employs the following notion of consistency, which I use to define a notion of "optimizing learnability".

Definition 4 *Given beliefs π^1, π^2, the belief-equivalent strategies f_2^1, f_1^2 and $\varepsilon \geq 0$, $\widehat{F} \subseteq F$ is* ε-consistent *if and only if, for each i, there exists an $f_i \in \widehat{F}_i$ and a full μ_{f^i}-measure set $A \in \mathscr{F}$ such that, for every $z \in A$ and all t, $f_i|_{z(t)}$ is an ε-best response to $f_{-i}^i|_{z(t)}$. Given π^1, π^2, f_2^1, f_1^2, $\varepsilon \geq 0$, an* ε-consistent *set $\widehat{F}$ and its ε-best responding element $f \in \widehat{F}$, the set $\widehat{F}$ then satisfies* (weak) optimizing learnability *if and only if, for all i, player i (weakly) learns to predict the path of play generated by f; $\widehat{F}$ satisfies* pure (weak) optimizing learnability *if and only if the set $\widehat{G}$ of pure strategies in $\widehat{F}$ satisfies (weak) optimizing learnability.*

The first part of this definition is just Nachbar's notion of consistency. Whilst a best response must also be a best response following any history with positive μ_{f^i}-measure, an ε-best response need not be an ε-best response in such a continuation, and indeed could be very suboptimal if the history receives low μ_{f^i}-measure. Nachbar's inconsistency rests on the suboptimality of a restricted strategy set on certain paths of play; if a player learns such a path, his continuation play will then fail to be ε-optimal, which is problematic since Nachbar requires any possible path of play to be learnable. By weakening learnability to require only that ε-optimal play is learnable, inconsistency can be avoided in many games, since such problematic continuations need no longer receive positive μ_{f^i}-measure. Intuitively, the "evil twin" strategies that drive Nachbar's (2005) inconsistency result are not ε-optimal in all of the games to which the result applies; hence, whilst Nachbar's learnability requires these strategies to be believed possible, *optimizing* learnability can give them zero μ_{f^i}-measure. In other words, the player will never need to learn "evil twin" paths of play under rational learning, and thus there is no implied failure of ε-optimization in continuation play.

Now, player 1's *minmax* payoff (for instance) is

$$\phi_1 = \min_{\alpha_2 \in \Delta(\Sigma_2)} \max_{\alpha_1 \in \Delta(\Sigma_1)} u_1(\alpha_1, \alpha_2).$$

Player 1's *pure action maxmin* payoff is

$$\Phi_1 = \max_{a_1 \in \Sigma_1} \min_{a_2 \in \Sigma_2} u_1(a_1, a_2).$$

Definition 5 (Nachbar 1997) *The stage game satisfies* MM *if and only if, for each player i, the pure action maxmin payoff is strictly less than the minmax payoff, $\Phi_i < \phi_i$.*

Matching pennies, rock–scissors–paper, battle of the sexes, and many coordination games satisfy MM. It is this broad class of games to which Nachbar's (2005) inconsistency result applies, but the inconsistent class of games is narrowed under Definition 4's notion of optimizing learnability, as I show in Proposition 1.

Given an action $a_1 \in \Sigma_1$ of player 1, let $\mathrm{BR}_2(a_1)$ denote the set of player 2's stage-game best responses to a_1; player 1's *pure action maxmin** payoff is then

$$\Phi_1^* = \max_{a_1 \in \Sigma_1} \min_{a_2 \in \mathrm{BR}_2(a_1)} u_1(a_1, a_2).$$

Definition 6 *The stage game satisfies* MM* *if and only if, for each player i, the pure action maxmin* payoff is strictly less than the minmax payoff, $\Phi_i^* < \phi_i$.*

The following result captures how Nachbar's inconsistency is altered when his learnability is replaced by optimizing learnability.

Proposition 1 *Suppose that MM* holds. Then there is a $\bar{\lambda} \in (0,1]$ such that, for any $\lambda_1, \lambda_2 \in [0, \bar{\lambda})$, there is an $\varepsilon > 0$ such that, for any $\widehat{F} \subseteq F$ and any beliefs, if $\widehat{F}$ is ε-consistent and satisfies CSP, then $\widehat{F}$ does not satisfy pure weak optimizing learnability.*

The proof of this result is relegated to Appendix A. Informally, if MM* holds, the players are not too patient and the strategy set satisfies optimizing learnability, then player 1 learns to predict that his payoff average is at most Φ_1^*, which is strictly less than his minmax payoff ϕ_1. However, the players of an ε-consistent strategy vector always expect to earn at least $\phi_i - \varepsilon'$ on average, in any continuation game with positive μ_f-measure, yielding a contradiction.

MM* is a much narrower class of games than MM; whilst matching pennies and rock–scissors–paper still belong to MM*, battle of the sexes and coordination games in general do not. That it should be difficult to learn the mixed-strategy equilibria of games such as matching pennies and rock–scissors–paper, but not necessarily the pure-strategy equilibria of coordination games, is in accord with what we would expect from the evolutionary game theory literature, and indeed with the problematic mixed-strategy equilibrium examples of Foster and Young (2001). Moreover, as we will see in the next section, it is possible to learn equilibrium in games outside of the class MM*. Note also that—by contrast with Nachbar's (2005) inconsistency result for the class MM—Proposition 1 imposes a maximum discount factor, further limiting its scope. The maximum discount factor creates a degree of myopia in the players' optimization that is crucial for the inconsistency result, as in Nachbar's result for his widest class of games NWD; Proposition 1 would not hold for the general discount factor in Nachbar's class MM result. Thus, when we replace Nachbar's learnability with optimizing learnability, we derive inconsistency for fewer games and fewer discount factors.

An apparently puzzling example in light of Proposition 1 is provided by stochastic fictitious play, which can be cast as a special case of Bayesian learning with i.i.d. strategies, and which converges to approximate equilibrium in repeated matching pennies. The restriction to i.i.d. strategies satisfies both CSP and weak learnability, but for games in the class MM optimal play is not i.i.d., and for games in the class MM* this is true even along the optimizing path of play. Hence, stochastic fictitious play fails to be ε-consistent in this case.

Proposition 1 is suggestive that the worst of Bayesian learning's convergence problems lie in games with interior mixed-strategy equilibria. That the latter are fundamentally problematic remains a consequence of Nachbar's inconsistency, as the following result shows. I will say that the strategy vector $f \in F$ is *fully mixed* if $f|_h(a) > 0$ for all $h \in H$ and all $a \in \Sigma$.

Corollary 1 *Suppose that MM holds. For any $\lambda_1, \lambda_2 \in [0, 1)$ there is an $\varepsilon > 0$ such that, for any $\widehat{F} \subseteq F$ and any beliefs π^1, π^2, if $\widehat{F}$ is ε-consistent and satisfies CSP, and any ε-optimal strategy vector $f \in \widehat{F}$ is fully mixed, then $\widehat{F}$ does not satisfy pure weak optimizing learnability.*

This follows from the second part of Nachbar's (2005) Theorem 1, since optimizing learnability implies Nachbar's learnability under a fully mixed f.

4 Learning while Optimizing

But why should we be interested in my notion of optimizing learnability? Because, as I show in this section, together with ε-consistency this notion is sufficient for convergence of Bayesian learning to approximate Nash equilibrium. So whilst there remains an inherent tension between learning and optimization in the class MM* of games, outside of this class learning and optimization can be reconciled. I then offer examples of such reconciliation in coordination games under 'dumb' and 'smart' beliefs.[4]

Given beliefs π^1, π^2, the belief-equivalent strategies f^1, f^2 and strategies $f = (f_1, f_2)$ *play eventually weakly ζ-like an ε-Nash equilibrium* if there exists a set $A \in \mathscr{F}$ such that:

i. $\mu_f(A) = 1$; and

ii. for every $z \in A$, there exists an $\bar{n}$ such that for all $n \geq \bar{n}$, $f|_{z(n)}$ and $f^i|_{z(n)}$, $i = 1, 2$, play weakly ζ-like an ε-Nash equilibrium.

[4] The term 'smart' here follows the usage of Noguchi (2015b), rather than Stahl (1993).

Theorem 1 *Let f be a behavior strategy vector, π^1, π^2 the beliefs of the players and f^1, f^2 the belief-equivalent strategy vectors. For any $\varepsilon > 0$, if f is $\varepsilon/2$-consistent and satisfies optimizing learnability, then f and f^1, f^2 play eventually weakly ε-like an ε-Nash equilibrium.*

The proof of this result is relegated to Appendix B.

Theorem 1 establishes the joint sufficiency of consistency and optimizing learnability for Bayesian learning of Nash equilibrium in repeated games. Thus, the tension established by Nachbar (1997, 2005) between consistency and learnability need pose no problem for Bayesian learning as long as optimizing learnability can be established. Proposition 1 establishes that this tension carries over to optimizing learnability in the class of games MM*. A key issue for the theorem is then the existence of a strategy vector and beliefs satisfying its conditions in games belonging to MM but not MM*. Below I offer two examples of such strategy vectors and beliefs in repeated coordination games.

'Dumb' learning under bounded memory Suppose that the stage game is a coordination game where $\Sigma_1 = \Sigma_2$ and each $a \in \Sigma$ is a Nash equilibrium if and only if $a_1 = a_2$. Given a t-length history $h = (z_1(t), z_2(t))$, let its *transpose* $\tau(h) = (z_2(t), z_1(t))$ be the history obtained by swapping the players associated with each action. A strategy vector $f \in F$ is *symmetric* if, for all $h \in H$, $f_1|_h = f_2|_{\tau(h)}$.

A natural strategy set satisfying Nachbar's (2005) CSP condition is the set $F^\kappa \subset F$, $\kappa \in \mathbb{N}$, of strategy vectors that have memory at most κ in the sense that, for all $f \in F^\kappa$, all $h \in H_\kappa$ and all $h', h'' \in H$, $f(h'h) = f(h''h)$. If both players know that strategies lie in F^κ, then the support of each belief π^i is a subspace of F^κ_{-i} (although Aumann's π^i-equivalent strategy vector f^i need not belong to F^κ).[5] I will refer to the resulting game as one of *bounded memory.* We know from Nachbar (1997) that a best response to a belief with full support on F^κ_{-i} will lie outside of F^κ_{-i} along some play paths. But I now show that F^κ may nonetheless satisfy $\varepsilon/2$-consistency and optimizing learnability.

Proposition 2 *Suppose that the stage game is a coordination game with $\Sigma_1 = \Sigma_2$, and that each $a \in \Sigma$ is a Nash equilibrium if and only if $a_1 = a_2$. Then there exists $\bar{\kappa} \in \mathbb{N}$ sufficiently high that, if each player i has a uniform prior belief $\tilde{\pi}^i$ on $F^{\bar{\kappa}}_{-i}$, then $F^{\bar{\kappa}}$ is $\varepsilon/2$-consistent and satisfies optimizing learnability.*

Proof. Given $\tilde{\pi}^i$, $\tilde{f}^i$ is the belief-equivalent strategy vector. Given any $\varepsilon' > 0$ and any $\kappa \in \mathbb{N}$, consider the restriction $\tilde{\pi}_{\varepsilon'}$ of $\tilde{\pi}$ to an ε'-neighborhood Y of the opponent's strategy

[5] The *support* of π^i is the largest closed subset of F_{-i} for which every open neighborhood of every point of the set has positive π^i-measure.

f_{-i} in F^{κ}_{-i}, and the probability measures $\tilde{\mu}$ and $\tilde{\mu}_{\varepsilon'}$ that $\tilde{\pi}$ and $\tilde{\pi}_{\varepsilon'}$ induce on the cylinder sets. $\tilde{\mu}_{\varepsilon'}$ is absolutely continuous with respect to $\tilde{\mu}$, and hence for any $\varepsilon'' > 0$ and $\tilde{\mu}_{\varepsilon'}$-almost every z, there exists an $\bar{n} \in \mathbb{N}$ such that $f|_{z(n)}$ plays weakly $(\varepsilon' + \varepsilon'')$-like $\tilde{f}^i|_{z(n)}$ for all $n \geq \bar{n}$ by the Blackwell and Dubins (1962) theorem.[6] To extend this statement to μ_f-almost every z, let $\widehat{A} \in \mathscr{F}$ be a set of full μ_f-measure. Since $\tilde{\pi}_{\varepsilon'}$ puts strictly positive measure on every open subset of Y, $\tilde{\mu}_{\varepsilon'}$ puts strictly positive probability on every (cylinder) subset of $\widehat{A}$—i.e. on every history possible under f. Thus, the players learn to predict the path of play generated by f.

Given $\varepsilon > 0$, fix each $\bar{\kappa} \in \mathbb{N}$ sufficiently high that, for each player i, $F^{\bar{\kappa}}_i$ contains a strategy f^*_i such that $f^*_i|_h$ is an $\varepsilon/3$-best response to $\tilde{f}_{-i}|_h$ for every $h \in H$ with $\mu_{\tilde{f}^i}(h) > 0$. Let $\hat{f} \in F^{\bar{\kappa}}$ be a symmetric strategy vector that plays a pure action profile in any period following the play of a symmetric action profile, and a full mixture of actions otherwise. There exists for $\mu_{\tilde{f}^i}$-almost every h some maximum probability $\bar{p}(h) > 0$ that a mixture of f^*_i and $\hat{f}_i$ may place on $\hat{f}_i|_h$ and remain an $\varepsilon/2$-best response to $\tilde{f}_{-i}|_h$. Let $\bar{f} \in F^{\bar{\kappa}}$ be a mixture of f^*_i and $\hat{f}_i$ such that $\bar{f}|_h$ places probability $\bar{p}(h)$ on $\hat{f}|_h$ for all such h; I claim that there exists an $\underline{n} \in \mathbb{N}$ such that $\bar{p}(z(n)) = 1$ for all $n \geq \underline{n}$ and $\mu_{\tilde{f}^i}$-almost every z. Supposing otherwise, for any $\underline{n} \in \mathbb{N}$, there must exist an $n \geq \underline{n}$ and a z with $\mu_{\tilde{f}^i}(z(n)) > 0$ such that $\bar{p}(z(n)) = 1$ is inconsistent with an $\varepsilon/2$-best response. But symmetric play is absorbing under $\hat{f}$, and with probability 1 a symmetric action profile is eventually played, after which each subsequent action profile is symmetric. Thus, the path of play induced by $\hat{f}$ is eventually symmetric, and hence $\hat{f}$ is eventually a Nash equilibrium. Moreover, if the value of $(\varepsilon' + \varepsilon'')$ above is $\varepsilon/4$, then there exists an $\bar{n} \in \mathbb{N}$ such that, for all $n \geq \bar{n}$, $\sup_{A \in \mathscr{F}_n} |\mu_{\bar{f}|_{z(n)}}(A) - \mu_{\tilde{f}|_{z(n)}}(A)| \leq \varepsilon/2$ by weak $(\varepsilon' + \varepsilon'')$-closeness of $\mu_{\bar{f}|_{z(n)}}$ and $\mu_{\tilde{f}|_{z(n)}}$. Hence, if $\bar{p}(z(n)) = 1$, $\bar{f}_i|_{z(n)}$ is an $\varepsilon/2$-best response to $\tilde{f}_{-i}|_{z(n)}$ for each i (by the normalization of payoffs to lie in the unit interval, and the continuity and per-period scale of U_i), contradicting the supposition and establishing the claim.

Note that there are paths with zero $\mu_{\bar{f}}$-measure that do not yield eventual symmetric play, but which can be generated by other elements of $F^{\bar{\kappa}}$; Nachbar's weak learnability requires such paths to be weakly learnable, but $\bar{f}$ does not eventually induce $\varepsilon/2$-best responses along them, and hence does not establish $\varepsilon/2$-consistency of $F^{\bar{\kappa}}$ under weak learnability. ■

Thus, consistency and optimizing learnability may be reconciled for the CSP strategy set $F^{\bar{\kappa}}$ in games belonging to MM but not MM*. It follows by Theorem 1 that the strategies and beliefs employed in the proof of Proposition 2 will eventually play weakly ε-like an ε-

[6] Recall that μ is *absolutely continuous with respect to* $\tilde{\mu}$ if $\mu(A) > 0$ implies $\tilde{\mu}(A) > 0$ for every $A \in \mathscr{F}$.

Nash equilibrium. Intuitively, consider strategies that eventually yield symmetric play with probability one, and the restriction of uniform priors on each $F^{\bar{\kappa}}_{-i}$ to an ε'-neighborhood of those strategies. This restriction approximates a Nash equilibrium, and since the restriction is absolutely continuous with respect to the original uniform prior, the latter must also eventually approximate a Nash equilibrium.[7] Given sufficiently high $\bar{\kappa}$, there must then exist approximately optimal strategies that place increasing weight on such symmetric strategies. Learnability would not be achievable in this way if we required it to hold along any possible play path, as opposed to along just the realized play path; for instance, paths with persistent asymmetric play would not yield eventual Nash equilibrium, but they also could not be generated by approximately optimal play, and hence pose a problem for Nachbar's learnability but not for optimizing learnability.

It is essential for the result that the players know not only that strategies have finite memory, but that there is a known bound on that memory. This is because the argument relies on the perpetual closeness of a strategy with respect to every strategy in its ε'-neighborhood in $F^{\bar{\kappa}}_{-i}$. If the players did not know a bound on the memory of strategies, then the ε'-neighborhood would be in the superspace of all finite-memory strategy vectors; hence it would be an infinite-dimensional neighborhood, with no uniform bound on the future differences in play of members of the neighborhood, i.e. no limit on the novelty arising in an open set of strategies in the infinite future.

As an instance of his characterization of the learnability of a set of probability measures, Noguchi (2015b, p. 425) shows that there exists a prior belief that leads its holder to learn to predict any finitely complex strategy vector. Using this characterization, Noguchi (2015a) proves that there exist prior beliefs that guarantee weak convergence of Bayesian learning to approximate Nash equilibrium under smooth near-optimal behavior and a complexity condition; in particular, prior beliefs in his model are carefully chosen to incorporate a fictitious-play-like learning procedure with "random search and (statistical) testing" in the style of Foster and Young (2003). In effect, a hypothesis-testing procedure is rendered rational in his model by the anticipation of its consequences in the players' prior beliefs.

In general, Noguchi (2015b, p. 431) views his learnability characterization as enabling us "to find out various types of 'smart' prior beliefs. . . [that lead] the player to learn to pre-

[7] Effectively, we have a reversal of Sandroni's (1998) approximate absolute continuity idea: beliefs merge with a strategy vector that forever weakly approximates the true one, rather than forever weakly approximating a strategy vector that merges with the true one. Lehrer and Smorodinsky's (1996) result that "diffusion" around a probability measure implies almost weak merging is not quite sufficient for the learning part of the result; the freedom that almost weak merging allows in a finite set of periods is unsuited to the continuity of utility and finite-game approximations on which Lemmas 6 and 7 rely. This is also the reason for the condition that players (not just weakly) learn the path of play in Theorem 1.

dict as many strategies of her opponents as possible". Proposition 2 shows that, in the case of such coordination games with bounded memory, a 'dumb' diffuse prior is sufficient to guarantee learnability of strategies, which moreover may be approximately optimal along the realized path of play. Note the manner in which this result escapes Nachbar's inconsistency: There are asymmetric histories inducing asymmetric, sub-optimal play, but these have zero μ_f-measure, and eventually low μ_{f^i}-measure. Thus, whilst not $\varepsilon/2$-consistent under Nachbar's weak learnability, F^κ is $\varepsilon/2$-consistent under optimizing learnability.

'Smart' learning by "principled players" A second example satisfying the conditions of Theorem 1 is provided by Sandroni (2000) in repeated coordination games with two actions and a single "cooperative" outcome that yields the highest payoffs for both players. Here, to avoid a further proliferation of notation, I informally outline an instance of Sandroni's result, and refer the reader to the original paper for a formal presentation of the general case.

Consider the game in Figure 1, which belongs to Sandroni's class of games, and also to MM, but not to MM*. Sandroni's Proposition 1 shows that there exist discount factors and beliefs (that believe in "strict reciprocity" in the opponent's play) under which it is optimal for the players always to "cooperate" by playing (H, T), and such that they are eventually confident that their opponent will also cooperate. The belief-equivalent strategy for a belief in strict reciprocity is of bounded complexity (Kalai and Stanford 1988), and hence belongs to a CSP strategy set $\widehat{F}$; since the optimal response of "always cooperate" has zero complexity, it is also contained in $\widehat{F}$, which thus satisfies consistency. Moreover, because the players learn to predict the path of play generated by the optimal strategies, $\widehat{F}$ satisfies optimizing learnability and play eventually approximates a Nash equilibrium.

Sandroni's result essentially provides 'smart' prior beliefs that allow rational learning of equilibrium play in his class of coordination games; his "principled players" adopt beliefs that anticipate the properties of optimal play in those games, in a seemingly natural way that improves upon the sort of 'dumb' belief (a diffuse prior, for instance) that would be required to learn *any* possible play. This 'smart' learning is not possible for fully mixed equilibria; Nachbar's inconsistency applies even to optimizing learnability in games such as repeated matching pennies.[8] However, the class of such problematic games is much smaller than those without a stage-game weakly dominant action.

[8] Noguchi (2015a) provides 'smart' prior beliefs that guarantee weak convergence of Bayesian learning to approximate Nash equilibrium (under smooth near-optimal behavior and a complexity condition) for any perturbed finite stage game, but the presence of a payoff perturbation means that this result is not a counterexample to the inconsistency of Bayesian learning in games such as repeated matching pennies.

5 Conclusion

Nachbar's (1997, 2005) inconsistency dealt a heavy blow to the enterprise of rational learning in repeated games, seemingly confining its scope to stage games of little strategic interest. More recently, positive results such as those of Noguchi (2015a, 2015b) have indicated that such a pessimistic conclusion is premature. This paper has located room for manoeuvre away from Nachbar's impossibility via a natural modification in his notion of learnability, requiring that only optimizing paths of play be learnable rather than any path of play. Such "optimizing learnability" is sufficient for Bayesian learning of Nash equilibrium, and is consistent with the restriction of beliefs to interesting strategy sets in a broad class of games that includes coordination games. The case that remains problematic is Bayesian learning of fully mixed equilibria, as suggested by Foster and Young (2001). But the results presented above show that Bayesian learning of pure-strategy equilibria in strategically interesting games is quite possible.

Appendix

A Proof of Proposition 1

In order to prove Proposition 1, I will need the following key concept.

Definition 7 *Fix* $\lambda_1, \lambda_2 \in [0,1)$ *and* $\varepsilon' > 0$. *Then* $\widehat{F} \subseteq F$ *has the* ε'-evil genius property *if and only if, for any* $g \in \widehat{F}$ *and any beliefs* f_2^1, f_1^2 *for which both players weakly learn to predict the path of play generated by* g, *there exists an* $h \in H$ *with* $\mu_g(C(h)) > 0$ *such that, if* $g_2|_h$ *is an* ε'*-best response to* $f_1^2|_h$, *then* $g_1|_h$ *is not an* ε'*-best response to* $f_2^1|_h$, *and similarly for player* 2. *In this case, each of* g_1 *and* g_2 *is said to be an* ε'-evil genius *against the other.*

The following is then immediate, and analogous to Nachbar's (2005) Theorem 2.

Lemma 1 *For any* $\lambda_1, \lambda_2 \in [0,1)$, *any* $\varepsilon' > 0$ *and any beliefs, if a set of pure strategies is* ε'*-consistent and has the* ε'*-evil genius property, then it does not satisfy weak optimizing learnability.*

The following is a simple variant of Nachbar's (2005) Lemma 2.

Lemma 2 *Consider any $\lambda_1 \in [0,1)$, any belief f_2^1, any $\varepsilon \geq 0$, any path $z \in \Sigma^\infty$, any behavior strategy f_1 such that $f_1|_{z(t)}$ is an ε-best response to $f_2^1|_{z(t)}$ for all t, and any $\iota > 0$. Consider any pure strategy g_1 such that, for any history $h \in H$, if $g_1(h) = a_1 \in \Sigma_1$, then $f_1|_h(a_1) > \iota$. Then $g_1|_{z(t)}$ is an $\varepsilon/(1-\lambda_1)\iota$-best response to $f_2^1|_{z(t)}$ for all t. A similar statement holds for player 2.*

Together with the following result (*cf.* Nachbar's Theorem 3), condition P, Lemma 1 and Lemma 4 below, it implies Proposition 1.

Proposition 3 *Suppose that MM* holds. Then there is an $\varepsilon' > 0$ and a $\bar{\lambda} \in (0,1]$ such that, for any $\lambda_1, \lambda_2 \in [0, \bar{\lambda})$, if $\widehat{F} \subseteq F$ satisfies CS, then $\widehat{F}$ has the ε'-evil genius property.*

Proof. Define $a_2^* : \Sigma_1 \to \Sigma_2$ by

$$a_2^*(a_1) = \arg \min_{a_2 \in \mathrm{BR}_2(a_1)} u_1(a_1, a_2).$$

If the right-hand side is not single-valued, an arbitrary selection may be made to be $a_2^*(a_1)$. The function a_1^* is defined similarly. Given $g_1 \in \widehat{F}_1$, define $G_2^*(g_1) \subset G_2$ to be the set consisting of all g_2 for which there exists a set $\mathbb{N}^\diamond \subseteq \mathbb{N}$ with $\lim_{n\to\infty} |\{1,\ldots,n\} \cap \mathbb{N}^\diamond|/n = 1$ such that, for all $n \in \mathbb{N}^\diamond$, letting z denote the path of play generated by (g_1, g_2) and letting $h = z(n)$,

$$g_2(h) = a_2^*(g_1(h)).$$

The definition of $G_1^*(g_2)$ is analogous.

The following lemma confirms that, for the class MM* of games, elements of $G_2^*(g_1)$ are evil geniuses against g_1. Informally, if MM* holds, then $g_1(h)$ is not a stage game ε'-best response to $a_2^*(g_1(h))$ for a density 1 set of periods, for ε' sufficiently small. Fix $\bar{\lambda}$ low enough to ensure that, for any $g \in \widehat{F}$ and any beliefs f_2^1, f_1^2 for which both players weakly learn to predict the path of play generated by g, an ε'-best response in any $h \in H$ with $\mu_g(C(h)) > 0$ must involve infinite repetition of stage-game best responses.

Lemma 3 *Suppose that MM* holds. Then there is an $\varepsilon' > 0$ such that, for any $\lambda_1, \lambda_2 \in [0, \bar{\lambda})$ and any pure strategy $g_1 \in G_1$, if $g_2 \in G_2^*(g_1)$, then g_2 is an ε'-evil genius against g_1. An analogous statement holds for any $g_2 \in G_2$.*

The proof of this result is just that of the second part of Nachbar's (2005) Lemma 3, with Φ_1^* in place of his M_1.

The following is a trivial modification of Nachbar's (2005) Lemma 4.

Lemma 4 *Suppose that a set $\widehat{G}$ of pure strategies satisfies CS, and consider any $g_1 \in \widehat{G}_1$. There is a $g_2 \in \widehat{G}_2 \cap G_2^*(g_1)$. An analogous statement holds with the roles of the players reversed.*

The result is then immediate from Lemmas 3 and 4. ■

B Proof of Theorem 1

In order to prove Theorem 1, I will first offer a number of constituent results. A vector of strategies, f, is a *weak ξ-subjective η-equilibrium* if there is a pair of (supporting) strategy vectors f^1, f^2 such that for each player i:

i. $f_i^i = f_i$;

ii. f_i is a ξ-best response to f_{-i}^i; and

iii. f plays weakly η-like f^i.

Lemma 5 *Let $f = (f_1, f_2)$ be a weak ψ-subjective 0-equilibrium of a finitely repeated game. There is a ψ-Nash equilibrium $\hat{f} = (\hat{f}_1, \hat{f}_2)$ that plays weakly 0-like f.*

This is a trivial modification of Fudenberg and Levine's (1993) Theorem 4. Intuitively, if f^1, f^2 are f's supporting strategy vectors, for each player $i \in \{1, 2\}$ change the play of each player $j \neq i$ to that given by f_j^i following each history that can be reached if i unilaterally deviates from f; the resulting modified strategy vector must play weakly 0-like a ψ-Nash equilibrium.

Lemma 6 *In finitely repeated games, for every $\theta > 0$ there is $\hat{\eta} > 0$ such that for all $\eta < \hat{\eta}$, if f is a weak ψ-subjective η-equilibrium, then there exists $\hat{f}$ such that:*

i. f plays weakly θ-like $\hat{f}$; and

ii. $\hat{f}$ is a ψ-Nash equilibrium.

Proof. Suppose to the contrary that there is $\theta > 0$ and a sequence of strategy vectors $f(m)$ such that (i) $f(m)$ is a weak ψ-subjective η_m-equilibrium, where $\eta_m \to 0$ as $m \to \infty$, and (ii) $f(m)$ does not play weakly θ-like any ψ-Nash equilibrium. Since $f(m)$ is a weak ψ-subjective η_m-equilibrium, there is a matrix $(f(m)_j^i)$ which sustains it. In finitely repeated games, each player has a finite number of pure strategies, so the set of behavior strategies is sequentially compact. Thus, without loss of generality, the sequences $\{f(m)\}_m$ and

$\{(f(m)^i_j)\}_m$ are converging (in the product topology) to, say, f and (f^i_j). As the utility functions are continuous in the product topology metrized by d, f is a weak ψ-subjective 0-equilibrium sustained by (f^i_j). Moreover, if η_m is close enough to zero, $f(m)$ plays weakly θ-like f. Using Lemma 5, we can find a ψ-Nash equilibrium $\hat{f}$ which plays weakly 0-like f. Thus, if η_m is sufficiently small, $f(m)$ plays weakly θ-like $\hat{f}$, which is a ψ-Nash equilibrium; a contradiction. ■

This is Kalai and Lehrer's (1993b) Remark 2 adapted to the product topology.

Lemma 7 (Kalai and Lehrer 1993b) *In infinitely repeated games, for every $\varepsilon > 0$ there is $\hat{\eta} > 0$ such that for all $\eta \leq \hat{\eta}$, if f is a weak ξ-subjective η-equilibrium, then there exists $\hat{f}$ such that:*

i. f plays weakly $\varepsilon/2$-like $\hat{f}$; and

ii. $\hat{f}$ is a $(\xi + \varepsilon/2)$-Nash equilibrium.

Proof. Let $\varepsilon > 0$. Observe first that there is an integer $l = l(\varepsilon)$ such that: (i) if a strategy k^l_i is a ψ-best response to k^l_{-i} in the l-fold repeated game, then any strategy k_i of the infinitely repeated game whose l-truncation coincides with k^l_i is a $(\psi + \varepsilon/4)$-best response to any k_{-i}, whose l-truncation coincides with k^l_{-i}; and (ii) if k_i is a ξ-best response to k_{-i} in the infinitely repeated game, then k^l_i is a $(\xi + \varepsilon/4)$-best response to k^l_{-i}.

Letting $\theta = \varepsilon/4$, there exists an $\hat{\eta}$ such that the conclusions of Lemma 6 hold for the l-fold repeated game. Let f be a weak ξ-subjective η-equilibrium for some $\eta < \hat{\eta}$. f^l is therefore a weak $(\xi + \varepsilon/4)$-subjective η-equilibrium in the l-fold repeated game. Therefore, by Lemma 6, it plays weakly $\varepsilon/4$-like some $(\xi + \varepsilon/4)$-Nash equilibrium, say $\hat{f}^l$.

To conclude the proof I need to define a strategy vector $\hat{f}$ of the infinitely repeated game whose l-truncation coincides with $\hat{f}^l$ and, moreover, have f play weakly $\varepsilon/2$-like it. Thus, I need only define $\hat{f}$ on histories longer than l. Let $h \in H_{l'}$, $l' > l$, be such a history; define $\hat{f}_i(h) = f_i(h)$. The argument in Kalai and Lehrer's (1993b) Theorem 1 proof then applies unchanged to establish that f plays weakly $\varepsilon/2$-like $\hat{f}$.

Recall that $\hat{f}^l$ is a $(\xi + \varepsilon/4)$-Nash equilibrium in the l-fold repeated game. Therefore, $\hat{f}$ is a $(\xi + \varepsilon/4 + \varepsilon/4)$-Nash equilibrium in the infinitely repeated game. ■

This is the general-ξ case of Kalai and Lehrer's (1993b) Theorem 1, the proof of which it follows closely. Loosely, starting with a weak ξ-subjective η-equilibrium f, consider its truncation to the finitely repeated game of length l. If l is large, then the truncated f is a weak ψ-subjective η-equilibrium of the finite game for some $\psi > \xi$. Moreover, by Lemma 6 it must approximately play weakly like some ψ-Nash equilibrium $\hat{f}$ of the finite game. I extend $\hat{f}$ to the infinite game by making it coincide with f after all histories longer than l. This extension makes f play close to $\hat{f}$ in the infinite game, and exploiting again the fact that l is large, $\hat{f}$ must be a $(\xi+\varepsilon/2)$-Nash equilibrium of the infinite game.

Proof of Theorem 1. Given $\varepsilon > 0$, fix $\eta \leq \varepsilon/2$ to be at most the value $\hat{\eta}$ in Lemma 7. By optimizing learnability, the players learn to predict the path of play, so that for μ_f-almost any path of play z there exists an $n(z) \in \mathbb{N}$ such that $f|_{z(n)}$ plays weakly η-like $f^i|_{z(n)}$ for all $n \geq n(z)$. It follows by $\varepsilon/2$-consistency of f that, for μ_f-almost every z and all $n \geq n(z)$, $f|_{z(n)}$ is an $\varepsilon/2$-subjective η-equilibrium. By Lemma 7, for all such n there exists an $(\varepsilon/2 + \varepsilon/2)$-Nash equilibrium $\hat{f}$ that plays weakly $\varepsilon/2$-like $f|_{z(n)}$, from which the result follows. ■

References

Aumann, R. J. 1964. “Mixed and Behavior Strategies in Infinite Extensive Games.” In *Annals of Mathematics Studies 52,* edited by M. Dresher, L. S. Shapley, and A. W. Tucker, 627–650. Princeton, NJ: Princeton University Press.

Blackwell, D., and L. Dubins. 1962. “Merging of Opinions with Increasing Information.” *Annals of Mathematical Statistics* 33:882–886.

Esponda, I., and D. Pouzo. 2016. “Berk–Nash Equilibrium: A Framework for Modeling Agents with Misspecified Models.” *Econometrica* 84:1093–1130.

Foster, D. P., and H. P. Young. 2001. “On the Impossibility of Predicting the Behavior of Rational Agents.” *Proceedings of the National Academy of Sciences of the USA* 98 (22): 12848–12853.

———. 2003. “Learning, Hypothesis Testing, and Nash Equilibrium.” *Games and Economic Behavior* 45:73–96.

Fudenberg, D., and D. K. Levine. 1993. “Self-Confirming Equilibrium.” *Econometrica* 61:523–545.

Jordan, J. S. 1995. “Bayesian Learning in Repeated Games.” *Games and Economic Behavior* 9:8–20.

Kalai, E., and E. Lehrer. 1993a. “Rational Learning Leads to Nash Equilibrium.” *Econometrica* 61:1019–1045.

Kalai, E., and E. Lehrer. 1993b. “Subjective Equilibrium in Repeated Games.” *Econometrica* 61:1231–1240.

———. 1994. “Weak and Strong Merging of Opinions.” *Journal of Mathematical Economics* 23:73–86.

Kalai, E., and W. Stanford. 1988. “Finite Rationality and Interpersonal Complexity in Repeated Games.” *Econometrica* 56:397–410.

Kuhn, H. W. 1953. “Extensive Games and the Problem of Information.” In *Contributions to the Theory of Games II,* Annals of Mathematics Study 28*,* edited by H. W. Kuhn and A. W. Tucker, 193–216. Princeton, NJ: Princeton University Press.

Lehrer, E., and R. Smorodinsky. 1996. “Compatible Measures and Merging.” *Mathematics of Operations Research* 21:697–706.

Levy, J. 2015. “Limits to Rational Learning.” *Journal of Economic Theory* 160:1–23.

Miller, R. I., and C. W. Sanchirico. 1997. “Almost Everybody Disagrees Almost All the Time: The Genericity of Weakly Merging Nowhere.” Columbia University Discussion Paper No. 9697–25.

———. 1999. “The Role of Absolute Continuity in ‘Merging of Opinions’ and ‘Rational Learning’.” *Games and Economic Behavior* 29:170–190.

Nachbar, J. H. 1997. “Prediction, Optimization, and Learning in Repeated Games.” *Econometrica* 65:275–309.

———. 2005. “Beliefs in Repeated Games.” *Econometrica* 73:459–480.

Noguchi, Y. 2015a. “Bayesian Learning, Smooth Approximate Optimal Behavior, and Convergence to ε-Nash Equilibrium.” *Econometrica* 83:353–373.

———. 2015b. “Merging with a Set of Probability Measures: A Characterization.” *Theoretical Economics* 10:411–444.

Nyarko, Y. 1998. “Bayesian Learning and Convergence to Nash Equilibria without Common Priors.” *Economic Theory* 11:643–655.

Sandholm, W. H. 2010. *Population Games and Evolutionary Dynamics.* The MIT Press.

Sandroni, A. 1998. “Necessary and Sufficient Conditions for Convergence to Nash Equilibrium: The Almost Absolute Continuity Hypothesis.” *Games and Economic Behavior* 22:121–147.

———. 2000. “Reciprocity and Cooperation in Repeated Coordination Games: The Principled-Player Approach.” *Games and Economic Behavior* 32:157–182.

Savage, L. J. 1954. *Foundations of Statistics.* New York: Wiley.

Stahl, D. O. 1993. “Evolution of Smart$_n$ Players.” *Games and Economic Behavior* 5:604–617.