Available online at www.sciencedirect.com

SCIENCE DIRECT®

Games and Economic Behavior 47 (2004) 157–171

GAMES and Economic Behavior

www.elsevier.com/locate/geb

Belief-based equilibrium ☆

Alvaro Sandroni [a,*] and Rann Smorodinsky [b]

[a] *Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA*

[b] *Davidson Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel*

Received 27 December 2001

Abstract

We introduce a new solution concept for short-sighted players engaging in a repeated interaction: a Belief-based equilibrium (BBE). In a BBE, players optimize myopically given their beliefs which are not necessarily correct, but are not contradicted by the data. We show that, if the stage game has a unique correlated equilibrium then the play of a BBE resembles a Nash equilibrium play. However, a BBE may not be a Nash equilibrium. In particular, in a BBE players may play deterministically when the only Nash equilibrium is in mixed strategies.

JEL classification: D83

Keywords: Equilibrium; Calibration; Rationality; Learning

---

1. Introduction

Assume players interact repeatedly in the same game and are known to act rationally and myopically. In a Nash equilibrium all players take best responses to correct beliefs. However, it may be impossible to infer from the observed data if a player is or is not randomizing. This motivates us to consider a new notion of equilibrium, a *Belief-based equilibrium* (BBE), which is based on the following assumptions:

☆ This is a revision of a manuscript previously entitled "Dynamic Foundations of Equilibrium."

[*] Corresponding author.

E-mail addresses: sandroni@kellogg.northwestern.edu (A. Sandroni), rann@ie.technion.ac.il (R. Smorodinsky).

doi:10.1016/S0899-8256(03)00152-0

(a) in every stage, each player forms a belief about what his opponents' actions will be in the next stage;
(b) all players take an action (pure or mixed) that provides them with the best expected outcome given their belief; and
(c) the observed data does not contradict players' beliefs, vis-à-vis a pre-specified set of calibration tests.

In a calibration test, players compare the long run frequencies of opponents' actions with the long run average belief and expect the frequencies to become closer to the long run average of their expectations. Such tests appear natural to people when they want to check the validity of their belief over a random outcome (Camerer, 1995). In our paper we assume that players also assume that when beliefs are correct that long run frequencies computed is subsequences, e.g., on odd periods, become close to long run averages of expectations. In fact, players expect this to occur simultaneously in many subsequences.

The notion of a BBE combines, in fact, three different strands of the literature: first, the Bayesian literature, where players use Bayes rule to update their strategies; second, the literature on calibration, culminating in the works of Forster and Vohra (1997, 1998) and Lehrer (1997, 2000); third, the literature that deals with bounded rationality and automata (e.g., Neyman, 1985; Rubinstein, 1986; Abreu and Rubinstein, 1988). Whereas in the latter branch of the literature players' strategies must be implementable by some finite automata or Turing machine we model players that can test their beliefs' validity with tests implementable by finite automata or Turing machines.

1.1. BBE and Nash equilibrium

It turns out that the notions of a BBE and Nash equilibrium are related. A straightforward application of the strong law of large numbers implies that, if players are playing a Nash equilibrium, they must be playing a BBE as well (Dawid, 1982).

Next, consider games with a unique correlated equilibrium, which is therefore the unique Nash equilibrium. We show that if players play a BBE then the frequencies of any finite sequence of action profiles equal the probabilities of those finite sequences implied by a repeated Nash equilibrium. We call this property *Nash normality*. Therefore, if players play a BBE, then an outsider, who applies calibration test, analogous to those applied by the players, will not reject the hypothesis that the players are playing a Nash equilibrium.

1.2. BBE and learning

In a companion paper, "Calibration with many checking rules" (Sandroni et al., 2003), we consider a forecaster who has no prior knowledge about the sequence of data that will be revealed. After observing the past data, the forecaster makes a prediction for the next period. We show that there exists a forecasting rule which ensures that the forecasts will be calibrated for *every* possible infinite string of data. Based on this result, we show a universal learning scheme which leads into a BBE. This learning scheme requires that players know their own payoffs but the player need not have any knowledge over the opponents' payoffs or their rationality.

There is no forecasting rule which ensures that the forecasts will be correct for all sequences of outcomes. In order to show convergence to correct beliefs, some assumptions are made on the prior belief (e.g., Blackwell and Dubins, 1962; Kalai and Lehrer, 1994; Lehrer and Smorodinsky, 1996; Sandroni, 1998). This shows that assuming calibrated beliefs is indeed weaker than assuming correct beliefs. We, furthermore, note some 'no-regret' qualities of the suggested learning scheme. Namely, players also do not regret, *à la* Hannan (Hannan, 1957), past actions.

Although the introduction first discusses BBE and then Nash normality, our exposition is reversed. We first define Nash normality and provide related results and then define BBE and discuss related results. In the next section we use some examples to motivate the definitions and results of the paper. Section 3 provides the basic model and Section 4 introduces Nash normality. Sections 5 and 6 provide the core definition of BBE and the main results. Section 7 provides a universal learning scheme leading to BBE.

2. Examples

An outsider observes the play path of myopic players and must determine whether or not the play is consistent with an independent sequence of Nash equilibrium plays. There are several ways in which the outsider could compare the play path with a Nash equilibrium. One of them is to check whether the empirical frequencies of stage game action profiles converge to the probability distribution determined by the Nash equilibrium. This criteria is common in the learning literature (see, for example, Foster and Vohra, 1998), but the following trivial example shows that it is a weak criterion.

Example 1. Consider the standard game of matching pennies:

	A	B
A	$(+1, -1)$	$(-1, +1)$
B	$(-1, +1)$	$(+1, -1)$

Assume that player 1 plays A for two periods and then B for two periods, repeatedly. Player 2 always alternates between A and B. The play path follows the cycle (AA), (AB), (BA), (BB), (AA), (AB), (BA), $(BB)\dots$. The empirical frequencies of actions profiles converge to the unique equilibrium distribution of 0.25 for each action profile. However, the play path does not resemble the Nash equilibrium play in which every action profile is played with probability 0.25. Indeed, the empirical frequency of (AA) followed by (BB) is zero, as opposed to the long-run anticipated average of $(0.25)^2$, when a Nash equilibrium is played independently over time.

Example 1 shows that the comparison between empirical frequencies and an independent sequence of Nash equilibrium plays' distribution should not be restricted to stage-game action profiles. This is so because the independent sequence of Nash equilibrium plays also pins down frequencies of all finite histories. An outsider could reasonably reject the hypothesis that the players are playing a Nash equilibrium, if the empirical frequency of *any* finite history (such as (AA), (AB) followed by (BB), (AA) three consecutive times, etc.) does not match the equilibrium distribution. This motivates us to consider the possibility that the outsider would use a richer set of tests than the comparison between the empirical frequencies of stage game action profiles and the Nash equilibrium distribution in the entire sequence.

We say that a play path is *Nash-normal* if the empirical frequency of *any* finite history matches an equilibrium distribution.[1] In other words, a play path that is *not* Nash-normal is *inconsistent* with a Nash equilibrium in the sense that for some finite history, the empirical frequency of its occurrence does not match the equilibrium distribution.

A direct consequence of the strong law of large numbers is that if players indeed play according to the same Nash equilibrium, independently over time, then the play path must be Nash-normal (with probability one). However, the following example shows that Nash-normal play paths may result even if players are not playing a Nash equilibrium.

Example 2. Let $\dot{s}$ be a concatenation of all the natural numbers in their dyadic expansion (e.g., $0 = 0$, $1 = 1$, $2 = 10$, $3 = 11$, $4 = 100$, $5 = 101$, etc). That is, $\dot{s} = \{0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, \ldots\}$. The sequence $\dot{s}$ is known in the number-theoretic literature as the Chempernowne number (Chempernowne, 1933). The relative frequency of any finite sequence of 0's and 1's of length r in the Chempernowne number is 0.5^r. Let $\dot{s}_t$ denote the tth digit in the Chempernowne number. Consider the game of matching pennies of Example 1. Assume that, at period t, player 1 plays A whenever $\dot{s}_{2t} = 1$and B otherwise. Player 2 plays A whenever $\dot{s}_{2t+1} = 1$ and B otherwise. A deterministic play cannot be a Nash equilibrium in matching pennies, but this play path is Nash-normal. To see this, note that (A, A) is played when $\dot{s}_{2t} = 1$ and $\dot{s}_{2t+1} = 0$. The frequency of 1's in the Chempernowne number is 0.5. The frequency of 0's in the periods in which the last outcome was 1 is also 0.5. So, the frequency of 10 and (A, A) is a quarter. Analogously, the frequency of any sequence of outcomes of length r is 0.25^r.

Example 2 motivates us to seek general conditions that ensure that the play paths are Nash-normal. We will maintain the standard assumption that, given their beliefs, players take best responses, but we will relax (considerably) the extreme assumption that players' beliefs are correct.

Example 3. Consider the game and strategies of Example 2, where players both believe their opponent randomizes equally between the two actions, then the long run frequencies converge to the average belief. In fact, this still holds in subsequences such as "all periods after (A, A) was played."

Example 3 does not describe a Nash equilibrium because players' beliefs are simply incorrect (players are not randomizing as they believe). However, we do not expect players to change their beliefs. This is the motivation for the new notion of equilibrium we introduce.

---

[1] We borrow the term *normal* from number theory. It refers to numbers such that the set of digits of their infinite expansion seem to be coming out of a uniform distribution (see Shiryayev, 1996, p. 394).

3. The basic model

3.1. The stage game

Let $N = \{1, 2, \ldots, n\}$ be the set of players. For any $i \in N$, let S_i be player i's (finite) action set. Let $S = \prod_{i=1}^{n} S_i$ be the set of all action profiles and let $S_{-i} = \prod_{j \neq i} S_j$ be the set of action profiles of i's opponents.

For each player $i \in N$, there exists a payoff function $u_i : S \to R$. The tuple $G = (S; u_1, \ldots, u_n)$ is the stage game. A *correlated equilibrium* for G is a probability measure, $\bar{\mu}$, over S such that $E_{\bar{\mu}}(u_i(s_i, s_{-i}) \mid s_i) \geqslant E_{\bar{\mu}}(u_i(s_i', s_{-i}) \mid s_i)$ for any $i \in N$, any $s_i' \in S_i$, and any $s_i \in S_i$ played with $\bar{\mu}$-strictly positive probability, where $E_{\bar{\mu}}(\cdot)$ is the expectation operator associated with $\bar{\mu}$. If the marginal distributions of $\bar{\mu}$ over the actions sets S_i are independent then $\bar{\mu}$ is a *Nash equilibrium*.

3.2. The repeated game

We denote by G^∞ the repeated game. Let S^t be the set of histories of length t. Let $\overline{S} \equiv \bigcup_{t=0}^{\infty} S^t$ be the set of all finite histories and let S^∞ be the set of all play paths. Given $s \in S^\infty$, we denote by s_t the tth coordinate of s and by $s^t = (s_1, s_2, \ldots, s_t) \in S^t$ the prefix of length t of s.

A strategy for player i, in the repeated game, is a function $\sigma_i : \overline{S} \to \Delta(S_i)$, where $\Delta(S_i)$ is the set of probability measures over S_i. We denote by Σ_i the set of strategies for i. Let $\Sigma = \prod_{i=1}^{n} \Sigma_i$ be the set of strategy profiles. Let $\sigma \in \Sigma$ be the actual strategy profile adopted by the players. Let λ_σ be the probability measure over S^∞ associated with σ.

Player i's belief is a function $b_i : \overline{S} \to \Delta(S)$. We assume that players know their own actions, i.e., the marginal distribution of $b_i(s^t)$ over S_i is equal $\sigma_i(s^t)$ (we say that b_i is *compatible* with σ_i). We also assume that players take myopic best responses to their beliefs, i.e., $E_{b_i(s^t)}(u_i) \geqslant E_\mu(u_i)$ for any measure $\mu \in \Delta(S)$, which has the same marginal distribution over S_{-i} as $b_i(s^t)$.

Given an action $x \in S_i$, let $b_i^x : \overline{S} \to \Delta(S)$ be the function with the property that the marginal of $b_i^x(s^t)$ over S_{-i} is identical to the marginal of $b_i(s^t)$ over S_{-i} and the marginal of $b_i^x(s^t)$ over S_i assigns probability one to x. So, $b_i^x(s^t)$ is player i's belief over outcomes, if player i plays x (regardless of whether or not x is optimal).

4. Nash-normal play paths

Definition 1. A function $C : \overline{S} \to \{0, 1\}$ is called a (*history-based*) *checking rule*.

Given a play path, $s \in S^\infty$ and a (history-based) checking rule, there is a subsequence of stages t for which $C(s^t) = 1$. We say that checking occurs in this subsequence. For example, a checking rule could always check, it could check only in odd periods, it could check when the last outcome was a specific action profile, etc.

For any $s \in S^\infty$, let $I_t(s)$ denote the $|S|$-dimension vector with 1 in the s_t-th coordinate and zero elsewhere.

Definition 2. The *calibration score* of s with respect to a probability measure μ on S and the checking rule C, at time t, is the following $|S|$-dimensional vector:

$$\phi_t(s, C, \mu) = \frac{\sum_{t=0}^{T} C(s^t) I_{t+1}(s)}{\sum_{t=0}^{T} C(s^t)} - \mu.$$

The calibration score is the difference between the probability measure and the empirical frequencies of outcomes, in the periods that follow checking.

Definition 3. The probability measure $\mu \in \Delta(S)$ *calibrates a checking rule* C *on* $s \in S^\infty$, if $\sum_{t=1}^{\infty} C(s^t)$ is finite or if $\lim_{t\to\infty} \phi_t(s, C, \mu) = \bar{0}$ (where $\bar{0}$ is the $|S|$-dimensional vector of zeroes).

A probability measure μ calibrates a rule on a path if the long-run empirical frequencies of outcomes, in the specified periods, are identical to μ.

Definition 4. For any set of checking rules M, we say that μ *calibrates* M *on* $s \in S^\infty$, if it calibrates all checking rules C in M on s.

Definition 5. Let M be a set of checking rules. A play path, $s \in S^\infty$, is M*-Nash-normal*, if there exists a Nash equilibrium of the stage game, $\bar{\mu}$, that calibrates M on s.

A play path is M-Nash-normal, if the empirical frequencies of outcomes coincide with Nash equilibrium of the stage game in all subsequences specified by the checking rules in M.

A Markovian checking rule of length r checks whenever the last r outcomes coincide with a pre-specified sequence of outcomes of length r. For any natural number r, let $M(r)$ be the set of all Markovian checking rules of length r. Let $\overline{M} \equiv \bigcup_{r=0}^{\infty} M(r)$ be the set of all Markovian checking rules. If a play path is $\overline{M}$-Nash-normal then we simply say it is *Nash-normal*. It is immediate to see that a play path is Nash-normal if and only if the frequency of any finite history of outcomes $(s_1, s_2, \ldots, s_t) \in S^t$ is eventually equal to $\prod_{j=1}^{t} \bar{\mu}(s_j)$, for some Nash equilibrium $\bar{\mu}$.

In the learning literature, it is common to require that the empirical frequencies of action profiles, in the stage game, converge to a Nash equilibrium probability distribution (see, for example, Foster and Vohra, 1998, and the literature on fictitious play). In the language of this paper, these play paths are $M(0)$-Nash-normal, where $M(0)$ contains only the checking rule that checks in every period. As demonstrated by Example 1, this criterion is weak. An $M(0)$-Nash-normal play path may not resemble a Nash equilibrium play. So, we look for stronger criteria of consistency between the play and the equilibrium. This can be obtained by requiring that the empirical frequencies match the Nash equilibrium not only in the entire sequence, but also in subsequences. These subsequences are determined by the checking rules in M. The larger the set of rules in M, the stronger the requirements a play path must satisfy to qualify as M-Nash-normal.

In a Nash-normal play path, the empirical frequency of any finite history of outcomes matches a Nash equilibrium distribution. However, other checking rules can be used to perform other tests as well. For example, in addition to the tests in $\overline{M}$, one could see if the empirical frequencies of action profiles match a Nash equilibrium distribution in the even periods, in the prime periods, etc. A large (but still countable) class of checking rules is $\widetilde{M}$, the class of all (history-based) checking rules which can be implemented by a recursive function or Turing machine.[2] We refer to a $\widetilde{M}$-Nash-normal play path as a *Turing–Nash-normal* play path. In a Turing–Nash-normal play path, empirical frequencies match a Nash equilibrium in all subsequences determined by a recursive function.

5. Belief-based equilibrium

Definition 6. Let C be a (history-based) checking rule and $x \in S_i$ be an action of player i. The (action-based) checking rule, C_i^x, for player i, is defined as follows:

$$C_i^x(s^{t+1}) = \begin{cases} 1 & \text{if } C(s^t) = 1 \text{ and } x = (s_{t+1})_i, \\ 0 & \text{otherwise.} \end{cases}$$

A history-based checking rule generates an action-based checking rule for each player i and action $x \in S_i$. This rule checks only when both conditions are satisfied; namely the original checking rule equals 1 after t stages and player i plays action x at time $t + 1$. It is worthwhile noting that for action-based checking rules we do not know, at period t, whether checking will occur at period t. We need to wait for player i's action at stage $t + 1$ to be realized.

Now we assume each player i compares his belief in the repeated game to the actual realization as follows:

Definition 7. A belief $b_i : \overline{S} \to \Delta(S)$ *calibrates an action-based checking rule* C_i^x on a play path $s \in S^\infty$, if $\sum_{t=0}^{\infty} C_i^x(s^{t+1})$ is finite or if

$$\lim_{T\to\infty} \frac{\sum_{t=0}^{T} C_i^x(s^{t+1})[I_{t+1}(s) - b_i^x(s^t)]}{\sum_{t=0}^{T} C_i^x(s^{t+1})} = \bar{0}.$$

Given a set of checking rules M, let $M_i \equiv \{C_i^x\colon\ C \in M,\ x \in S_i\}$ be the set of all action-based checking rules, for player i, associated with the checking rules in M.

Definition 8. A belief b_i *calibrates* M_i on a play path $s \in S^\infty$ if it calibrates all action-based checking rules in M_i on s. If player i's belief calibrates $\overline{M}_i$ on a play path $s \in S^\infty$, then we say that player i's *belief is calibrated on* s.

---

[2] This follows directly from the way Turing machines are represented, using a finite alphabet, a finite state space and a mapping from some finite space into {Left, Right, Stay}. For more details, see Hopcroft and Ullman (1979, Chapter 7).

A sophisticated version of the law of large numbers (Dawid, 1982) shows that if the beliefs are correct then they must also be calibrated.[3] So, if the forecasts are eventually not calibrated then they must be incorrect and the player should revise his beliefs.

5.1. Equilibrium concepts in the repeated game

Definition 9. A profile of beliefs and strategies $(b_i, \sigma_i)_{i\in N}$, is a (myopic) *M-BBE* if for all i:

(i) σ_i is player i's myopic best response to his beliefs and σ_i is compatible with b_i;
(ii) b_i calibrates M_i on λ_σ-almost every play path $s \in S^\infty$.

That is, an M-BBE is a set of beliefs which are calibrated on the play paths generated by myopic optimization (for all action-based checking rules associated with M). A $\overline{M}$-BBE is simply referred to as a BBE (recall that $\overline{M}$ is the set of all Markovian checking rules).

A (myopic) Nash-equilibrium of the repeated game is defined as in an M-BBE except that (ii) is replaced by the condition $b_i = \sigma$. In a Nash equilibrium, beliefs and actual strategies coincide. In a BBE, players' beliefs may not be correct, but they are calibrated.

Proposition 1. *Any* (*myopic*) *Nash equilibrium of the repeated game is a Belief-based equilibrium.*

Proof. A direct corollary from the main result in Dawid (1982).

The converse of Proposition 1 is not true. Consider the play in the matching pennies game in Example 2. Assume both players believe their opponent randomizes with equal probabilities. Beliefs and actual strategies do not coincide, but the play resembles the realization of a distribution identical to the player's beliefs. Beliefs and best responses are a BBE, but not a Nash equilibrium.

If an equilibrium is suppose to formalize an intuitive notion of stability, then there is no reason to assume that players' beliefs are correct. A player will only modify his forecasting method if he perceives it as flawed. An M-BBE makes explicit the tests players use to check their forecasting method. As long as the data does not contradict the forecasting record, in the players' view, players are assumed to maintain their forecasting method.

6. BBE and Nash normality

Proposition 2. *Given a history-based checking rule C, assume that players take myopic best responses to their beliefs and for all players $i \in N$ and actions $x \in S_i$, player i's belief b_i calibrates the action-based checking rule C_i^x on a play path $s \in S^\infty$. Also assume that*

---

[3] More precisely, given any countable set of checking rules M, if player i's beliefs are correct then b_i calibrates M_i on almost all play paths $s \in S^\infty$.

C checks infinitely often on s (i.e., $C(s^t) = 1$ infinitely often) and that for a subsequence of periods $T_1, T_2, \ldots$, where $T_k < T_{k+1}$, the limit of empirical frequencies

$$\eta \equiv \lim_{k\to\infty} \frac{\sum_{t=0}^{T_k} C(s^t) I_{t+1}(s)}{\sum_{t=0}^{T_k} C(s^t)} \quad \textit{exists.}$$

Then, η is a correlated equilibrium.

Proof. See Appendix A.

Foster and Vohra (1998) have demonstrated Proposition 2 for the special case that C is the checking rule active in every period. Proposition 2 is a generalization of their result for arbitrary checking rules. Provided that players' beliefs calibrate the action-based checking rules, Proposition 2 demonstrates that if the empirical frequencies converge, they converge to a correlated equilibrium (and not necessarily to a Nash equilibrium). In general, the convergence of empirical frequencies cannot be guaranteed (even in the special case that C is the always active checking rule). To see this, consider the standard coordination game:

$$\begin{array}{c|c|c|} & L & R \\ \hline T & (2,2) & (0,0) \\ \hline B & (0,0) & (1,1) \\ \hline \end{array}.$$

Consider the Nash equilibrium of the repeated game in which the play path is (TL) in the odd periods and (BR) in the even periods. Let C be the 'always active' checking rule. The frequencies of action profiles converge to (TL) and (BR) with equal probability. This distribution is a correlated equilibrium of the stage game, but not a Nash equilibrium. Moreover, in the odd periods, the empirical frequencies converge to a different correlated equilibrium ((TL) with probability one). So, Proposition 3 shows that, whenever the limit of the empirical frequencies exists, it will coincide with a correlated equilibrium, but not necessarily the same correlated equilibrium if there are many equilibria in the stage game. Note however that the limit of the empirical frequencies may not necessarily exist. To see this simply consider a play path that is either (TL) or (BR). In this play path players settle on one of these alternatives for a block of periods of length j and then move to the other alternative for a block of periods of length 10^j and so on and so forth. In this play path the frequency of (TL) does not converge.

We now show that if the stage game has a unique correlated equilibrium then, in an M-BBE, the empirical frequencies of action profiles converge to a Nash equilibrium distribution, on the subsequences of periods where checking occurs. In other words, the play path is M-Nash-normal.

Proposition 3. *Let G be a stage game with a unique correlated equilibrium. Let $(b_i, \sigma_i)_{i=1}^n$ be an M-belief-based equilibrium of G^∞. Then, λ_σ-almost surely, the play path is M-Nash-normal.*

Proof. See Appendix A.[4]

Corollary. *Let G be a stage game with a unique correlated equilibrium. Let $(b_i, \sigma_i)_{i=1}^{n}$ be a M-belief-based equilibrium of G^∞. Then, λ_σ-almost all play paths of a BBE are Nash-normal.*

Proposition 3 shows that, if the stage game has a unique correlated equilibrium then both the play paths of a (myopic) Nash equilibrium and the play paths of a BBE are Nash-normal. In this sense, the two concepts are observationally equivalent. So, a BBE can be used to provide foundations to a Nash equilibrium. Any learning model that leads to correct beliefs (i.e., where $b_i = \sigma$ for all i) will also lead to calibrated beliefs (see, for example, Kalai and Lehrer, 1993). However, in the models of Jordan (1991, 1995) and Nyarko (1994), beliefs are not eventually correct, but it is straightforward to show that, if the realized stage game has a unique correlated equilibrium, they are calibrated. There are many examples of ad-hoc learning rules that lead to $M(0)$-Nash-normal play paths, but not to correct beliefs (see, for example, Foster and Vohra, 1998; Hart and Mas-Colell, 2000; Brown, 1951; Fudenberg and Levine, 1995; Fudenberg and Kreps, 1993). It is relatively easy to demonstrate that any of these rules may be modified to lead to Nash-normal play paths.

The converse of Proposition 3 is not true. A play path can be Nash-normal even if beliefs and best responses are not a BBE. Therefore, in principle, one could obtain Nash-normal play paths under weaker assumptions than those required by a BBE. This is demonstrated in the example below.

Example 4. Consider the matching pennies game in Example 1. Let $\dot{s}_t$ denote the tth digit in the Chempernowne number. At period t, player 1 believes that player 2 plays A, whenever $\dot{s}_{2t} = 1$ and B otherwise. Player 2 believes that player 1 plays B whenever $\dot{s}_{2t+1} = 1$ and A otherwise. If both players play best replies, then player 1 plays A, whenever $\dot{s}_{2t} = 1$ and B otherwise. Similarly, player 2 believes that player 2 plays A with probability one. The data shows that, in these periods, player 2 plays A calibrated. In fact, player 1's beliefs do not even calibrate $M(0)_1$, where $M(0)$ contains only the always active checking rule.

7. Universal forecasting schemes

In this section, we show that players can ensure that their beliefs will be calibrated (but not that their beliefs will be eventually correct) without *any* knowledge about how the opponents will play. We show the existence of universal prediction schemes, i.e., forecasting schemes with the property that the beliefs are calibrated on *every* play path. This result is presented in greater detail in our companion paper Sandroni et al. (2003).

---

[4] The proof of Proposition 3 shows that this result holds path by path, i.e., the play paths such that player i's beliefs calibrate M_i, $i \in N$, are M-Nash-normal.

No standard belief, as defined in Section 3, is a universal forecasting scheme (Oakes, 1985). To construct a universal forecasting scheme, we use a random belief which maps past plays and histories of realized forecasts into probability distributions over next period beliefs. In what follows, the players draw their actual belief from the random belief and optimize their action accordingly. In particular, a player does not consider a random belief as a compounded lottery against which he will optimize.

We do not know of good decision-theoretic foundations for the assumption of random beliefs. So, this forecasting scheme is not intended to be a descriptive dynamic model leading to a Belief-based equilibrium. However, it does make an important distinction between calibrated and correct beliefs. It is possible to ensure that the beliefs are eventually calibrated, but it is not possible to ensure that the beliefs are eventually correct. In this sense, the assumption of calibrated beliefs is significantly weaker than the assumption of correct beliefs.

Although our motivation to use random beliefs in this section is technical we point out that the experimental economics literature provides strong evidence that players do not react to a random belief (decision under uncertainty) in the way they react to the derived compounded lottery (decision under risk), see Camerer (1995).

Definition 10. A random belief for player i, in the repeated game, is a function $\tilde{b}_i : \bigcup_{t=0}^{\infty} (\Delta(S) \times S)^t \to \Delta(\Delta(S))$.

At period t, the realization of $\tilde{b}_i$ induces a forecast over the opponents' play (the marginal of the realization of $\tilde{b}_i$ over S_{-i}). We assume that players take a pure best reply to these realized forecasts. So, the randomness in actions is only a consequence of the randomness in beliefs.

Definition 11. Given a set of checking rules M and a repeated play $s_{-i} \in (S_{-i})^{\infty}$ of i's opponents, the (standard) belief b_i is M-calibrated on s_{-i} if b_i calibrates M_i on the associated play path, $s \in S^{\infty}$, $s = (s^t, \ldots)$, where $(s_{t+1})_i$ is the best response for $b_i(s^t)$ and the opponents' actions coincide with s_{-i}, i.e., $(s_{-i})_t = (s_t)_{-i}$.

By Kolmogorov's Extension Theorem, any random belief $\tilde{b}_i$ determines a unique probability measure $\tilde{b}_i^*$ on the space Ω of beliefs b_i.

Definition 12. A random belief $\tilde{b}_i$ is an M-universal forecasting scheme if, given any repeated play, $s_{-i} \in (S_{-i})^{\infty}$ of i's opponents, $\tilde{b}_i^*$-almost-surely, the beliefs b_i are M-calibrated on s_{-i}. We say that an $\overline{M}$-universal forecasting scheme is a universal forecasting scheme.

By definition, if player i adopts an M-universal forecasting scheme, then, regardless of the actions of i's opponents, player i's beliefs are M-calibrated, with probability one according to i's randomization.

7.1. Existence

Proposition 4. *For any countable set of checking rules M, there exists an M-universal forecasting scheme $\tilde{b}_i$ for each player $i \in N$.*

Proof. Let $S_i = \{s_1, \ldots, s_K\}$ be player i's stage game action set. Let $\mathcal{D} = \{D_1, \ldots, D_K\}$ be a partition of $\Delta(S_{-i})$ such that player i chooses the action s_k whenever the next period forecast is in D_k. So, an action-based checking rule, for an action s_k of player i, is equivalent to a forecast-based checking rule based on the subset D_k, as defined in Sandroni et al. (2003). Proposition 4 is an immediate consequence of the main result in Sandroni et al. (2003) (applied to the state space of opponent's actions S_{-i}) and the assumption that player i correctly anticipates i's best reply. □

Proposition 4 is a generalization of the result of Foster and Vohra (1997), who focus on the special case where $M = M(0)$, i.e., the only checking rule is the always active checking rule. The sets $\overline{M}$ and $\widetilde{M}$ are countable sets. Therefore, a direct corollary of Proposition 4 is that there exists a universal forecasting scheme for player i which ensures that, regardless of the actions of i's opponents, player i's beliefs are calibrated (or even $\widetilde{M}$-calibrated).

A direct corollary of Propositions 3 and 4 is that if players use the universal forecasting scheme when repeatedly playing a game with a unique correlated equilibrium then taking a best response to their belief at each stage results in a Nash-normal play path, almost surely.

Remark. By Proposition 2, each player can construct a universal prediction scheme knowing only his own payoffs and the opponents' action set. Therefore, even if each player does not know the opponents' payoffs, they could generate beliefs using a universal prediction scheme and take myopic best responses to them. This would result in a BBE and, if the stage game has a unique correlated equilibrium, in a Nash-normal play path.

7.2. Regret

We now consider a thought experiment that demonstrates another application of Proposition 4: Assume that a decision maker can take two actions, T or B. Nature simultaneously selects between L and R. The decision makers' payoffs are given by the matrix below:

$$\begin{array}{c|c|c|}
 & L & R \\ \hline
T & 1 & 0 \\ \hline
B & 0 & 1 \\ \hline
\end{array}$$

Assume the decision maker plays according to a (myopic) best response to an M-universal forecasting scheme. Call this strategy sM. The decision maker selects T when he believes that nature will play L with probability greater or equal to 0.5 and B otherwise. Assume that $M = M(0)$, i.e., the only checking rule in M is the always active checking rule. By Proposition 4, the decision maker's beliefs are $M(0)$-calibrated. Hence, nature plays L more (or equally) often than nature plays R, in the periods that the decision maker selected T. Analogously, in the periods that the decision maker selected B, nature plays L less (or equally) often than R. Consider the alternative strategy that replaces B for T. That is, play B when $sM(0)$ plays T and play $sM(0)$ otherwise. The average payoff under this alternative strategy is not greater than under $sM(0)$ because in the periods that $sM(0)$ plays T and the alternative strategy plays B, the frequency of L is greater or equal to 0.5. So, for every sequence of nature moves, the average payoff obtained under $sM(0)$ is greater (or equal) than the average payoff for playing B (or T) in every period. This property of $sM(0)$ is sometimes called Hannan-consistency or no-regret property.

We now assume that the decision maker's strategy is, $s\widetilde{M}$, a best response to a $\widetilde{M}$-universal prediction scheme. Consider a new alternative strategy which is equal $s\widetilde{M}$ unless the $s\widetilde{M} = T$ and the stage t is an odd number. By a similar argument, we can show that the average payoff obtained by this alternative strategy cannot be greater than under $s\widetilde{M}$. A similar property holds if the replacement was made in stages t which are prime numbers, or any scheme that such replacements can be implemented by a Turing Machine.

The argument above shows that there exists a strategy, $s\widetilde{M}$, that generates higher (or equal) long-run average payoffs than any other strategy that replaces B for T (or T for B) in a way that can be implemented by a recursive algorithm. This result is quite general. The interested reader is referred to Lehrer (2000) who makes a formal presentation of this result and to related results by Blackwell (1954), Freund and Schapire (1999), Fudenberg and Levine (1999), Hannan (1957), Hart and Mas-Collel (2000, 2001), and Rustichini (1999). We simply note that the scheme presented in this section has good normative properties for long-run average payoffs.

Acknowledgments

We thank Uri Gneezy, Martin Dufwenberg, participants in several workshops, and an anonymous referee for valuable comments. Financial support from the NSF Grant SBR 9730385, BSF Grant 97-00113/1, and the Bergmann Memorial Research Grant is gratefully acknowledged. Sandroni also thanks financial support from the National Science Foundation Grant SES 0109650. All errors are ours.

Appendix A

Proof of Proposition 2. In order to show that η is a correlated equilibrium, it is sufficient to show that for any $x \in S_i$, played with η-strictly positive probability, $x \in BR(\eta_{-i} \mid s_i = x)$, where η_{-i} is the marginal distribution of η on S_{-i} and $BR(\eta_{-i} \mid s_i = x)$ is the set of actions which are best responses to the conditional distribution of η_{-i} given x.

Let $x \in S_i$ be an i's action played with η-strictly positive probability. By assumption,

$$\lim_{k\to\infty} \sum_{t=0}^{T_k} C\big(s^t\big) = \infty \quad \text{and} \quad \eta \equiv \lim_{k\to\infty} \frac{\sum_{t=0}^{T_k} C(s^t) I_{t+1}(s)}{\sum_{t=0}^{T_k} C(s^t)}.$$

Therefore,

$$\lim_{k\to\infty}\sum_{t=0}^{T_k} C_i^x\left(s^{t+1}\right)=\infty \quad \text{and} \quad \lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C_i^x(s^{t+1})I_{t+1}(s)}{\sum_{t=0}^{T_k} C_i^x(s^{t+1})} \quad \text{exist}$$

and are equal to the conditional distribution of η, x given.

By definition,

$$\lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C_i^x(s^{t+1})[I_{t+1}(s)-b_i^x(s^t)]}{\sum_{t=0}^{T_k} C_i^x(s^{t+1})}=0$$

$$\Rightarrow \quad \lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C_i^x(s^{t+1})b_i^x(s^t)}{\sum_{t=0}^{T_k} C_i^x(s^{t+1})} \quad \text{exists.}$$

Let $BR_i^{-1}(x)$ denote the set of stage-game beliefs over S_{-i} for which x is the best response. Let $B \equiv BR_i^{-1}(x) \times x$. Note that B is compact and convex. By definition, $b_i^x(\hat{s}^t) \in B$, and therefore,

$$\lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C_i^x(s^{t+1})b_i^x(s^t)}{\sum_{t=0}^{T_k} C_i^x(s^{t+1})} \in B \quad \Rightarrow \quad \lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C_i^x(s^{t+1})I_{t+1}(s)}{\sum_{t=0}^{T_k} C_i^x(s^{t+1})} \in B.$$

Equivalently, $x \in BR(\eta_{-i} \mid s_i = x)$. □

Proof of Proposition 3. Let $s \in S^\infty$ be a path such that, for all $i \in N$, b_i calibrates M_i. Let C be a (history-based) checking rule in M. Assume, without loss of generality, that $C(s^t)=1$ infinitely often (otherwise any probability measure on $\Delta(S)$ calibrates C on s). Let $T_1, T_2, \ldots$ be any subsequence of periods for which

$$\bar{\eta} \equiv \lim_{k\to\infty}\frac{\sum_{t=0}^{T_k} C(s^t)I_{t+1}(s)}{\sum_{t=0}^{T_k} C(s^t)} \quad \text{exists.}^5$$

That is, $\bar{\eta} \in \Delta(S)$ is an arbitrary limit point of the frequencies of play along C.

By Proposition 2, $\bar{\eta}$ is a correlated equilibrium of the stage game. However, the stage game is assumed to have a unique correlated equilibrium, which itself must be a Nash equilibrium. Therefore, all limits points of empirical frequencies along C are equal to the unique Nash equilibrium $\bar{\eta}$. So, the limit of empirical frequencies along C must exist and is equal to $\bar{\eta}$. □

References

Abreu, D., Rubinstein, A., 1988. The structure of Nash equilibrium in repeated games with finite automata. Econometrica 56, 993–1008.

Blackwell, D., 1954. Controlled random walks. In: Noordhoff, Erven P., Groningen, N.V. (Eds.), Proc. Int. Congress of Mathematicians, vol. 3. North-Holland, Amsterdam, pp. 335–338.

---

[5] The existence of η is ensured because $\Delta(S)$ is compact.

Blackwell, D., Dubins, L., 1962. Merging of opinions with increasing information. Ann. Math. Statist. 38, 882–886.

Brown, G., 1951. Iterative solution of games by fictitious play. In: Koopmans, Tjalling C. (Ed.), Activity Analysis of Production and Allocation. Wiley, New York.

Camerer, C., 1995. Individual decision-making. In Chapter 8 of: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton Univ. Press.

Chempernowne, D., 1933. The construction of a decimal normal in the scale of ten. J. London Math. Soc. 8, 254–260.

Dawid, A.P., 1982. The well calibrated Bayesian. J. Amer. Statist. Assoc. 77 379, 605–613.

Foster, D., Vohra, R., 1997. Calibrated learning and correlated equilibrium. Games Econ. Behav. 21, 40–55.

Foster, D., Vohra, R., 1998. Asymptotic calibration. Biometrika 85, 379–390.

Freund, Y., Schapire, R., 1999. Adaptive game using multiplicative weights. Games Econ. Behav. 29, 79–103.

Fudenberg, D., Kreps, D., 1993. Learning mixed equilibria. Games Econ. Behav. 5, 320–367.

Fudenberg, D., Levine, D., 1995. Universal consistency and cautious fictitious play. J. Econ. Dynam. Control 19, 1065–1089.

Fudenberg, D., Levine, D., 1999. Conditional universal consistency. Games Econ. Behav. 29, 104–130.

Hannan, J., 1957. Approximation to Bayes risk in repeated plays. In: Dresher, M., Tucker, A.W., Wolfe, P. (Eds.), Contributions to the Theory of Games. Princeton Univ. Press, pp. 97–139.

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

Hart, S., Mas-Colell, A., 2001. A general class of adaptative strategies. J. Econ. Theory 98, 26–54.

Hopcroft, J.E., Ullman, J.D., 1979. Introduction to Automata Theory, Languages and Computation. Addison–Wesley.

Jordan, J., 1991. Bayesian learning in normal form games. Games Econ. Behav. 3, 60–81.

Jordan, J., 1995. Bayesian learning in repeated games. Games Econ. Behav. 9, 8–20.

Kalai, E., Lehrer, E., 1993. Rational learning leads to Nash equilibrium. Econometrica 61, 1019–1045.

Kalai, E., Lehrer, E., 1994. Weak and strong merging of opinions. J. Math. Econ. 23, 73–100.

Lehrer, E., Smorodinsky, R., 1996. Compatible measures and merging. Math. Operations Res. 21-3, 697–706.

Lehrer, E., 1997. Any inspection rule is manipulable. Econometrica.

Lehrer, E., 2000. A wide range no-regret theorem. Games Econ. Behav.

Neyman, A., 1985. Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. Econ. Letters 19, 227–229.

Nyarko, Y., 1994. Bayesian learning in repeated games leads to correlated equilibria in normal form games. Econ. Theory 4, 821–841.

Oakes, D., 1985. Self-calibrated priors do not exist. J. Amer. Statist. Assoc. 80, 339.

Rubinstein, A., 1986. Finite automata play the repeated prisoners' dilemma. J. Econ. Theory 39, 83–96.

Rustichini, A., 1999. Minimizing regret: the general case. Games Econ. Behav. 29, 224–243.

Sandroni, A., 1998. Necessary and sufficient conditions for convergence to Nash equilibrium: The almost absolute continuity hypothesis. Games Econ. Behav. 22, 121–147.

Sandroni, A., Smorodinsky, R., Vohra, R., 2003. Calibration with many checking rules. Math. Operations Res. 28-1, 141–153.

Shiryayev, A.N., 1996. Probability, 2nd Edition. Springer.