10-1997

Calibrated Learning and Correlated Equilibrium

Dean P. Foster
University of Pennsylvania

Rakesh V. Vohra

Recommended Citation

Foster, D. P., & Vohra, R. V. (1997). Calibrated Learning and Correlated Equilibrium. *Games and Economic Behavior, 21* (1-2), 40-55. http://dx.doi.org/10.1006/game.1997.0595

Calibrated Learning and Correlated Equilibrium

Abstract
Suppose two players repeatedly meet each other to play a game where:

1. each uses a learning rule with the property that it is a calibrated forecast of the other's plays, and

2. each plays a myopic best response to this forecast distribution.

Then, the limit points of the sequence of plays are correlated equilibria. In fact, for each correlated equilibrium there is some calibrated learning rule that the players can use which results in their playing this correlated equilibrium in the limit. Thus, the statistical concept of a calibration is strongly related to the game theoretic concept of correlated equilibrium.

Disciplines
Behavioral Economics | Statistics and Probability

Calibrated Learning and Correlated Equilibrium

Dean P. Foster
University of Pennsylvania
Philadelphia, PA 19104 *

Rakesh V. Vohra
Ohio State University
Columbus OH 43210 †

First draft: May 1992, Revised: June 1993,
This version: October 17, 1996

*Email:foster@hellspark.wharton.upenn.edu

†Email:vohra.1@osu.edu

Abstract

Suppose two players meet each other in a repeated game where:

1. each uses a learning rule with the property that it is a calibrated forecast of the others plays, and
2. each plays a best response to this forecast distribution.

Then, the limit point of the sequence of plays are Correlated Equilibria. In fact, for each Correlated equilibrium there is some calibrated learning rule that the players can use which result in their playing this correlated equilibrium in the limit. Thus, the statistical concept of calibration is strongly related to the game theoretic concept of correlated equilibrium.

1 Introduction

The concept of a Nash Equilibrium (NE) is so important to game theory that an extensive literature devoted to its defense and advancement exists. Even so, there are aspects of the Nash equilibrium concept that are puzzling. One is why any player should assume that the other will play their Nash equilibrium strategy? Aumann (1987) says: "This is particularly perplexing when, as often happens, there are multiple equilibria; but it has considerable force even when the equilibrium is unique."

One resolution is to argue that the assumption about an opponent's plays are the outcome of some learning process (see for example Chapter 6 of Kreps (1991a)). Learning is modeled as recurrent updating. Players choose a best reply on the basis of their forecasts of their opponents future choices. Forecasts are described as a function of previous plays in the repeated game. Much attention has focused on developing forecast rules by which a Nash equilibrium (or its refinements) may be learned. Many rules have been proposed and convergence to Nash equilibrium has been established under certain conditions (see Skyrms 1990). For example, Fudenberg and Kreps (1991) introduce the class of rules satisfying a property called 'asymptotic myopic bayes.' They prove that *if* convergence takes place, it does so to a NE. Notice that convergence is not guaranteed. In summarizing other approaches, Kreps (1991b) points out, "in general convergence is not assured." This lack of convergence serves to lessen the importance of NE and its refinements.

On the positive side Milgrom and Roberts (1991) have shown that any learning rule that requires the player to make approximately best responses consistent with their expectations, play tends towards the serially undominated set of strategies. They call such learning rules adaptive and prove that if the sequence of plays converges to a NE (or correlated equilibrium) then each players play is consistent with adaptive learning.

Learning, as we have described it, takes place at the level of the individual. An important class of learning models involve learning at the level of populations (evolutionary models). Here the different strategies are represented by individuals in the population. In particular a mixed strategy would be represented by assigning an appropriate fraction of the population to each strategy. A pair of individuals is selected at random to play the game. Individuals do not update their strategies but their numbers wax and wane according to their average (suitably defined) payoff. Even in this environment convergence to a NE is not guaranteed. On the positive side, results analogous to Milgrom and Roberts have been obtained by Samuelson and Zheng (1992).

A second objection to NE is that it is inconsistent with the Bayesian perspective. A Bayesian player starts with a prior over what their opponent will select and chooses a best response to that. To argue that Bayesians should play the NE of the game is to insist that they each choose a *particular* prior. Aumann (1987) has gone further and argued that the solution concept consistent with the Bayesian perspective is not NE but Correlated Equilibrium (CE). Support for such a view can be found in Nau and McCardle (1990) who characterize CE in terms of the no arbitrage condition so beloved by Bayesians. Also, Kalai and Lehrer (1994) show that Bayesian players with uncontradicted beliefs learn a correlated equilibrium.

In this note, we provide a direct link between the Bayesian beliefs of players to the conclusion that they will play a CE. We do this by showing that a CE can be 'learned'. We do not particular a specific learning rule, rather, we restrict our attention to learning rules that possess an asymptotic property called calibration. The key result is that if players use *any* forecasting rule with the property of being *calibrated*, then, in repeated plays of the game, the limit points of the sequence of plays are correlated equilibria.

The game theoretic importance of calibration follows from a theorem of Dawid (1992). Given the Bayesians prior look at the forecasts generated by the posterior. The sequences of future events on which this forecast will not be calibrated, have measure zero. That is the Bayesian's prior assigns probability zero to such outcomes. Thus, under the common prior assumption, a bayesian would expect all the other players to be using their posterior, and hence to be calibrated. Now using our result that calibration implies correlated equilibria, and the common prior assumption shows that bayesians expect that in the limit, they will be playing a correlated equilibrium. This provides an alternative prove to Aumann's proof that the common prior assumption and rationality implies a correlated equilibrium. If the common prior assumption holds then it is common knowledge that all players are calibrated. If the players use a Bayesian forecasting scheme that is calibrated, then, by the above, in repeated plays of the game, the limit points of the sequence of plays are correlated equilibria.

In the next section of this paper we introduce notation and provide a rigorous definition of some of the terms used in the introduction. Subsequently we state and prove the main result of our paper. For ease of exposition we consider only the 2-person case. However, our results generalize easily to the n-person case.[1]

[1] See discussion after Theorem 3.

2 Notation and Definitions

For $i = 1, 2$, denote by $S(i)$ the finite set of pure strategies of player i and by $u_i(x, y) \in \Re$ the payoff to player i where $x \in S(1)$ and $y \in S(2)$. Let $m = |S(1)|$ and $n = |S(2)|$. A *correlated strategy* is a function h from a finite probability space ? into $S(1) \times S(2)$, i.e., $h = (h_1, h_2)$ is a random variable whose values are pairs of strategies, one from $S(1)$ and the other from $S(2)$. Note that if h is a correlated strategy, then $u_i(h_1, h_2)$, is a real valued random variable.

So as to understand the definition of a correlated equilibrium, imagine an umpire who announces to both players what ? and h are. Chance chooses an element $g \in$? and hands it to the umpire who computes $h(g)$. The umpire then reveals $h_i(g)$ to player i only and nothing more.

Definition: *A correlated strategy h is called a correlated equilibrium if:*

$$E(\ u_1(h_1, h_2)\ ) \geq E(\ u_1(\Phi(h_1), h_2)\ ) \textit{ for all } \Phi : S(1) \rightarrow S(1),$$

and,

$$E(\ u_2(h_1, h_2)\ ) \geq E(\ u_2(h_1, \Phi(h_2))\ ) \textit{ for all } \Phi : S(2) \rightarrow S(2),$$

Thus, a CE is achieved when no player can gain by deviating from the umpire's recommendation, assuming the other player will not deviate either. The deviations, are restricted to be functions Φ of h_i because player i knows only $h_i(g)$. For more on CE see Aumann (1974) and Aumann (1987).

We turn now to the notion of calibration. This is one of a number of criteria used to evaluate the reliability of a probability forecast. It has been argued by a number of writers (see Dawid (1982)) that calibration is an appealing minimal condition that any respectable probability forecast should satisfy. Dawid offers the following *intuitive* definition:

> Suppose that, in a long (conceptually infinite) sequence of weather forecasts, we look at all those days for which the forecast probability of precipitation was, say, close to some given value p and (assuming these form an infinite sequence) determine the long run proportion ρ of such days on which the forecast event (rain) in fact occurred. The plot of ρ against p is termed the forecaster's *empirical calibration curve*. If the curve is the diagonal $\rho = p$, the forecaster may be termed *well calibrated*.[2]

To give the notion a formal definition, suppose that player 1 is using a forecasting scheme f. The output of f in round t of play is an n-tuple $f(t) = \{p_1(t), \dots, p_n(t)\}$ where $p_j(t)$ is the forecasted probability that player 2 will play strategy $j \in S(2)$ at time t. Let $\chi(j,t) = 1$ if player 2 plays their j-th strategy in round t and zero otherwise. Denote by $N(p,t)$ the number of rounds up to the t-th round that f generated a vector of forecasts equal to p. Let $\rho(p,j,t)$ be the fraction of these rounds for which player 2 plays j, i.e.,

$$\rho(p,j,t) = \sum_{s=1}^{t} \frac{I_{f(s)=p}\chi(j,s)}{N(p,t)},$$

if $N(p,t) > 0$ and zero otherwise.

The forecast f is said to *calibrated with respect to the sequences of plays made by player 2* if:

$$\lim_{t\to\infty} \sum_{p} |\rho(p,j,t) - p_j| \frac{N(p,t)}{t} = 0$$

[2]Dawid (1982) page 605. His notation has been changed to match ours.

for all $j \in S(2)$. Notice that taking $0/0 = 0$ is now seen not to matter since the only time it will occur is if $N(p,t) = 0$, and thus it would be multiplied by zero anyway. Roughly, calibration says that the empirical frequencies *conditioned on the assessments* converge to the assessments. This is to be contrasted with the asymptotic myopic bayes condition of Fudenberg and Kreps which says that the empirical frequencies in round t converge together with the assessments in round t.

3 Calibration and Correlated Equilibrium

It is clear from the definition of correlated strategies that a CE is simply a joint distribution over $S(1) \times S(2)$ with a particular property. Hence, we focus on $D_t(x,y)$, the fraction of times up to time t that player 1 plays x and player 2 plays y. This is the empirical joint distribution. We assume that when players select their best response (for a given forecast) they use a a stationary and deterministic tie breaking rule; say the lowest indexed strategy.

Theorem 1 *Let π be the set of all correlated equilibria. If each player uses a forecast that is calibrated against the others sequence of plays, and then makes a best response to this forecast, then,*

$$\min_{D \in \pi} \max_{x \in S(1), y \in S(2)} |D_t(x,y) - D(x,y)| \to 0$$

as t, the number of rounds of play, tends to infinity.

PROOF: Observe first that the nm-tuple each of whose components is of the form $D_t(x,y)$ lies in the $nm-1$ dimensional unit simplex. By the

Bolzano-Weirstrass theorem any bounded sequence in it contains a convergent subsequence. Thus, for any subsequence $\{D_{t_i}(x, y)\}$ and $D(x, y)$ such that

$$\sum_{\substack{x \in S(1) \\ y \in S(2)}} |D_{t_i}(x, y) - D(x, y)| \to 0,$$

we need to show that D is a CE.

For each $x \in S(1)$ let $M_b(x)$ be the set of mixtures over $S(2)$ for which x is a best response. $M_b(x)$ is a closed convex subset of the $n-1$ dimensional simplex. Let $M_p(x)$ be the set of mixtures where player 1 actually plays x given that the forecast is in $M_p(x)$. By the assumption that players choose best responses, $M_p(x) \subseteq M_b(x)$. Further, $\{M_p(x) : x \in S(1)\}$ forms a partition of the simplex. The empirical conditional distribution of $y \in S(2)$ given that player 1 played x is $\frac{D_{t_i}(x,y)}{\sum_{c \in S(2)} D_{t_i}(x,c)}$. This converges to $\frac{D(x,y)}{\sum_{c \in S(2)} D(x,c)}$ as long as $\sum_{c \in S(2)} D_{t_i}(x, c)$ does not converge to zero. If it did, it would mean that the proportion of times that x is played tends to zero. Hence, in the limit, player 1 never plays x, so it can be ignored. To complete the proof it suffices to show that the n-tuple whose y-th component is $\frac{D(x,y)}{\sum_{c \in S(2)} D(x,c)}$ is contained in $M_b(x)$. Observe that:

$$\begin{aligned}
D_{t_i}(x, y) &= t_i^{-1} \sum_{r \le t_i : f(r) \in M_p(x)} \chi(y, r) \\
&= t_i^{-1} \sum_{p \in M_p(x)} \sum_{r \le t_i : f(r) = p} \chi(y, r) \\
&= t_i^{-1} \sum_{p \in M_p(x)} \rho(p, y, t_i) N(p, t_i) \\
&= t_i^{-1} \sum_{p \in M_p(x)} p_y N(p, t_i) + \\
&\quad + t_i^{-1} \sum_{p \in M_p(x)} (\rho(p, y, t_i) - p_y) N(p, t_i)
\end{aligned}$$

Since the forecasts being used are calibrated, the second term in the last expression goes to zero as t tends to infinity. Note:

$$\sum_{p \in M_p(x)} p_y \frac{N(p, t_i)}{\sum_{q \in M_p(x)} N(q, t_i)} \in M_b(x)$$

because it is a convex combination of vectors in $M_b(x)$ { recall, $M_p \subseteq M_b$}, and $M_b(x)$ is convex. Therefore

$$\frac{D(x, y)}{\sum_{c \in S(2)} D(x, c)} = \lim_{t_i \to \infty} \sum_{p \in M_p(x)} p_y \frac{N(p, t)}{\sum_{p \in M_p(x)} N(p, t)}$$

which is then the y^{th} component of a vector in $M_b(x)$ also.

We have shown that any sequence $\{D_{t_i}(x, y)\}$ contains a convergent subsequence whose limit is a CE. The theorem now follows. □

In some sense the result above is not surprising. We know from Milgrom and Roberts (1991) if players use best responses they eliminate dominated strategies. Secondly, the calibration requirement forces limit points to satisfy an additional equilibrium requirement. Correlation arises because players are able to condition on previous plays.

It is natural to ask if Theorem 1 would hold with a *non-stationary* tie-breaking rule. The following version of *matching pennies* shows that this is not possible. In each round the row player will forecast that there is a 50%

Matching Pennies

	h	t
H	1\-1	-1\1
T	-1\1	1\-1

chance that column will play heads and a 50% chance that column will play tails, i.e., (0.5, 0.5) is the forecast. The column player will do likewise. Given these forecasts there is a tie for the best reply. Consider the following tie breaking rule: on even numbered rounds play heads and tails on the other rounds. Notice that the resulting sequence of plays will be: Tt, Hh, Tt, Hh, ... Clearly the forecasts of each player are calibrated, but the distribution of plays does not converge to a CE.

Theorem 1 raises the question of how a calibrated forecast is to be produced. Oakes (1985), has shown that there is no *deterministic* forecast that is calibrated *for all possible* sequences of outcomes. Our requirements are more modest. Given a game, and a correlated equilibrium of this game, is there a sequence of plays and a deterministic forecasting rule depending only on observed histories that is calibrated? The next theorem provides a positive answer to this question.

Definition: *Call a point of the distribution $D(x, y)$ a* limit point of calibrated forecasts *if there exist deterministic best reply functions $R_i(\cdot)$ and calibrated forecasting rules p_i such that if each player i, plays $R_i(p_i)$, then the limiting joint distribution will be $D(x, y)$.*

Definition: *Let λ be the set of all distributions which are limit points of calibrated forecasts.*

Using this notation we can restate Theorem 1 as saying that $\lambda \subset \pi$.

We can represent every game by a vector in $\Re^{2mn}$, where each component corresponds to a players payoff. A set of games is of measure zero if the corresponding set of points in $\Re^{2mn}$ has Lebesgue measure zero.

Theorem 2 *For almost every game $\lambda(G) = \pi(G)$. In other words, for almost every game, the set of distributions which calibrated learning rules can converge to is identical to the set of correlated equilibriums.*

Proof: Because of Theorem 1 we need only prove that $\pi \subset \lambda$. Let (x_j, y_j) be a deterministic computable sequence such that the limiting joint distribution is $D(x, y)$. At time j, have player 1 forecast

$$p_{1,j}(\cdot) = D(x_j, \cdot) \Big/ \sum_{y \in S(2)} D(x_j, y)$$

and player two forecast

$$p_{2,j}(\cdot) = D(\cdot, y_j) \Big/ \sum_{x \in S(1)} D(x_j, y) \ .$$

By the assumption that the joint distribution converges to $D(x, y)$, it is clear that both of these forecasts are calibrated. Further, x_j is in fact a best response to the forecast $p_{1,j}(\cdot)$, and y_j to $p_{2,j}(\cdot)$. So, define $R_1(p)$ such that for all j, $x_j = R_1(p_{1,j})$ and similarly for $R_2(p)$. These forecasts and these best reply functions are the key idea of the proof. In fact, in the situation where $R_1(\cdot)$ and $R_2(\cdot)$ are both well defined we have completed the proof.

But, $R_1(\cdot)$ and $R_2(\cdot)$ might not be well defined. In other words, there might be two different strategies x' and x'' such that $x_{j'} = x'$ and $x_{j''} = x''$, then $p_{1,j'}(\cdot) = p_{1,j''}(\cdot) = p^*$. This is where the "almost every game" condition comes into play.

Almost every game has the property that all the sets $M_b(x)$ have non-empty interior. To see why this is the case, observe that $M_b(x)$ is formed by the intersection of half-spaces. Start with a closed convex set with non-empty interior, C, say and add these half-spaces one at a time. We can choose C

to be the simplex of all mixed strategies. Consider a half-space H, chosen at random such that the coefficients that define H are continuous with respect to lebesgue measure. We claim that the intersection of C and H is either the empty set, or a set with an open interior.

Pick a point p in the interior of C. Let q be the point in the boundary of H which is closest to p. Let v be the ray from p to q and d its length. Both v and d have continuous distributions since they are a continuous transformation of the half-space H. Now consider distribution of d conditional on v. Given v there is a unique d such that H will be tangent to C and not contain C . The conditional probability of d taking this value is 0. Hence the unconditional probability is zero also.

The interiors of the sets of the form $M_b(x)$ are disjoint. [3] Thus, near the point p^* there are points $p^{x'}$ and $p^{x''}$ such that the unique best response to $p^{x'}$ is x' and the unique best response to $p^{x''}$ is x''. Forecasting $p^{x'}$ or $p^{x''}$ instead of p^* makes the reply function well defined. Unfortunately, when the forecast of $p^{x'}$ is made, the actual frequency will turn out to be p^*. Thus, the calibration score will be off by at most $|p^{x'} - p^*|$. If we can choose $p^{x'}$ to be convergent to p^* solves this last problem and our proof is complete.

Define a sequence $p_i^{x'} = (1 - 1/i)p^* + (1/i)p^{x'}$. Then $p_i^{x'}$ converges to p^* and for all i, $p_i^{x'}$ has x' as its unique best reply. For each i forecast $p_i^{x'}$ sufficiently many times to ensure that there is a high probability that the empirical distribution is within $1/i$ of p^*. With high probability the empirical frequency conditional on forecast $p_i^{x'}$ will be within $2/i$ of $p_i^{x'}$ and hence the calibration score will converge to zero. □

[3] The interiors and the union of the boundaries would form a partition.

To see why theorem 2 only holds for almost every game and not every game, consider the following game:

Example of $\pi \neq \lambda$

	1	2	3
A	2\2	0\3	0\1
B	2\2	0\1	0\3
C	2\0	1\1	1\0

If ROW randomizes between A and B (with equal probability) and COL plays 1, then this is a Correlated Equilibrium with a payoff of (2,2). But, the only point in λ is the distribution which puts all its weight on point (C,2) which yields a payoff of (1,1). This is because: $M_b(A) = M_b(B) = \{(1,0,0)\}$ and $M_b(c)$ is the entire simplex. So, if $R_{\text{ROW}}((1,0,0)) = A$, then ROW will never play strategy B, and likewise if $R_{\text{ROW}}((1,0,0)) = B$, then ROW will never play A. So, a mixture of A and B is impossible and thus the payoff (2,2) is impossible. Thus, $\pi \neq \lambda$.

Can Theorem 1 be strengthened such that convergence to Nash Equilibrium is assured instead of to a CE? The previous theorem shows if one assumes only calibration, one gets *any* CE in π. So, without further assumptions on the forecasting rule, convergence to Nash cannot be assured. In particular by adding an assumption that the limit exists does not refine the equilibrium attained (in contrast with Fudenberg and Kreps who show that if a limit exists, it must be Nash). This is because Theorem 2 does not just find an accumulation point it finds a direct limit.

Is it easy to construct a forecast that is calibrated? Given the impossibility theorem of Oakes (1985) the existence of a *deterministic* scheme that is calibrated for *all* sequences is ruled out. However, a randomized forecasting scheme is possible.

Theorem 3 (Foster and Vohra 1991) *There exists a randomized forecast that player 1 can use such that no matter what learning rule player 2 uses, player 1 will be calibrated. That is to say, player* $1's$ *calibration score*

$$C_t \equiv \sum_p \sum_{j \in S(2)} |\rho(p,j,t) - p_j| \frac{N(p,t)}{t} \tag{1}$$

converges to zero in probability. In other words, for all $\epsilon 00$ *we have that* $\lim_{t\to\infty} P(C_t < \epsilon) = 1$.

Proof: See the appendix.

The important thing to notice about Theorem 3 is that each player can individually choose to be calibrated. The other player can not foil this choice. Player 1 does not have to assume that player 2 is using an exchangeable sequence, nor that the player 2 is rational. Player 1 is still calibrated if player 2 plays any arbitrary sequence. Secondly, the proof is constructive, i.e., there is an explicit algorithm for producing such a forecast.[4] To extend this result to the n-person case the forecasting rules must predict the joint distribution of what everyone else will play.

If in Theorem 1 we require only that the players use a forecasting rule that is close to calibrated in the sense of Theorem 3, we obtain:

Corollary *There exists a randomized forecasting scheme, such that if both player 1 and player 2 follow this scheme, then FOR ANY normal form matrix game and for all* $\epsilon > 0$*, there exists a* $t_0 > 0$*, such that for all* $t > t_0$*,*

$$P(\min_{D\in\pi} \max_{x,y} |D_t(x,y) - D(x,y)| < \epsilon) > 1 - \epsilon.$$

[4]The most involved step is inverting a matrix.

In other words, D_t converges in probability to the set π under the Hausdorff topology.

4 The Shapley Game and Fictitious Play

The most famous of learning rules for games is called Fictitious Play (FP), first conceived in 1949 by George Brown. In a two person game it goes as follows:

Definition: *Definition of Fictitious Play: Row computes the proportion of times up to the present that Column has played each of his/her strategies. Then, Row treats these proportions as the probabilities that Column will select from among his/her strategies. Row then selects the strategy that is his/her best response. Column does likewise.*

In 1951 Julia Robinson proved that FP converges to a NE in 2 person zero sum games.

After the Robinson paper, interest naturally turned to trying to generalize Robinson's theorem to non-zero sum games. In 1961, K. Miyasawa proved that FP converges to a NE in 2-person non-zero sum games where each player has at most two strategies.[5] However, in 1964 Lloyd Shapley dashed hopes of a generalization by describing a non-zero sum game consisting of three strategies for each player in which FP did not converge to a NE. In this section we show that FP doesn't converge to a Correlated Equilibrium. We use Shapley's original example:

[5] See Monderer and Shapley (1993) for other situations in which FP converges.

Payoff Matrix for Shapley Game

	1	2	3
1	1\0	0\1	0\0
2	0\0	1\0	0\1
3	0\1	0\0	1\0

As observed by Shapley, FP in this game will oscillate between 6 states, (1,1) then (1,2), then $(2,2),(2,3),(3,3),(3,1)$, then repeat. Fictitious play stays longer and longer in each state, so the periods of oscillation get larger and larger. There is only one Correlated equilibrium with support on these six states.[6] It assigns probability 1/6 to each state. Fictitious play is never close to this distribution.[7] Thus, it does not converge to a CE.

[6] Using Nau and McCardle (1990) the following linear program produces all the CE.

$$p_{11} \geq p_{12} \geq p_{22} \geq p_{23} \geq p_{33} \geq p_{31} \geq p_{11},$$

$$p_{13} \leq p_{11}, p_{13} \leq p_{23},$$

$$p_{21} \leq p_{22}, p_{21} \leq p_{31},$$

$$p_{32} \leq p_{33}, p_{32} \leq p_{12}.$$

Which is equivalent to the $LP : p_{11} = p_{12} = p_{22} = p_{23} = p_{33} = p_{31} = p_{11}, p_{13} \leq p_{11}, p_{21} \leq p_{11}, p_{32} \leq p_{11}$. Adding the constraint that $p_{13} = p_{21} = p_{32} = 0$, this LP has a unique solution of $p_{11} = p_{12} = p_{22} = p_{23} = p_{33} = p_{31} = 1/6$.

[7] This can be see either by direct calculation, or by the following trick. If Fictitious play was ever close to this CE, then the marginals would have to be close to $(1/3, 1/3, 1/3)$. But, these marginals correspond to the Nash Equilibrium. Shapley created this example precisely to show that the marginals didn't converge to the marginals of the Nash equilibrium, in fact the marginals are bounded away from the $(1/3, 1/3, 1/3)$ point. Thus the Nash equilibrium is not an accumulation point of the sequence of plays. Thus, we know that the marginals are never close to being correct, and thus the joint distribution is also never close.

The Shapley game is interesting because it has a CE which is not a mixture of Nash Equilibriums. [8] Theorem 3 tells us that there are calibrated learning rules which will then converge to this CE. The expected payoff is $(1/2, 1/2)$ which Pareto dominates the Nash payoff of $(1/3, 1/3)$.

Postscript

Earlier versions of this paper as well as presentations of the results at various conferences have generated a deal of follow on papers on calibration and its connections to game theory. In this section, we give a brief description of some of this work.

Theorem 3, which establishes the existence of randomized forecasting scheme that is calibrated has prompted a number of alternative proofs. The first of these was due to Sergiu Hart (personal communication) and is particularly simple and short. It makes use of the mini-max theorem. The draw back is that the scheme implied by the method is impractical to implement. Independently, Fudenberg and Levine (1995) also gave a proof using the minimax theorem. The approach is more elaborate than Hart's but produces a forecasting scheme that is practical to implement. In a follow up paper Fudenberg and Levine [1996] consider a refinement of the calibration idea that involves the classification of observations into various categories. For this refinement they derive a procedure that yields almost as high a time-average payoff as could be obtained if the player chooses knowing the conditional distribution of actions given categories. If players use such a procedure, long run the time average play resembeles a correlated equilibrium.

[8] The unique Nash Equilibrium for this game is $(1/3, 1/3, 1/3)$ vs $(1/3, 1/3, 1/3)$. So, any CE which isn't Nash, is also not a mixture of Nash Equilibriums.

Our own proof of Theorem 3 (which is described in the appendix) is based on establishing the existence of a forecasting scheme that has a property called no-regret. An proof along the lines of Theorem 1 shows that a no-regret procedure would also lead to a correlated equilibrium. Hart and Mas-Collel (1996) have extended this idea in many ways. First they proved a very elegant proof of no-regret based on Blackwell's (195?) vector mini-max theorem. Second they modify this scheme which requires a matrix inversion to one that involves regret-matching. This greatly reduces the computations required to implement the procedure. The simplified procedure no longer has the no-regret property but it will converge to a correlated equilibrium. Their theorem is much harder to prove since they can't simply appeal to a no-regret/calibration property as we have done.

Kalai, Lehrer and Smorodinsky (1996) have recently shown that the notion of calibration is mathematically equivalent to that of merging. This allows one to establish relationships between convergence results based on merging and those based on calibration and so derive some new convergence results.

Appendix

This appendix provides a telegraphic proof of Theorem 3. For more details see Foster and Vohra (1991).

We will first prove a property called "no-regret." Consider k forecasts each with a loss or penalty at time t of $0 \leq L_t^i \leq 1$ for $i = 1, \ldots, k$. Now consider a randomized forecast which picks forecast i at time t with probability w_t^i. We define the loss from using the combined forecast to be the weighted sum of the losses of each forecast, namely, $\sum_{i=1}^{k} w_t^i L_t^i$.

Definition: *The regret generated by changing all i forecasts to j forecasts is* $R_T^{i\to j} \equiv \max\{0, S_T^{ij}\} = S_T^{ij} I_{S_T^{ij}>0}$ *where* $I_{x>0}$ *is the indicator function and*

$$S_T^{ij} \equiv \sum_{t=1}^{T} w_t^i (L_t^i - L_t^j).$$

We choose the probability vector w_t so that it satisfies the following flow conservation equations:

$$(\forall i) \qquad w_t^i \sum_{j=1}^{k} R_{t-1}^{i\to j} = \sum_{j=1}^{k} w_t^j R_{t-1}^{j\to i}.$$

The duality theorem of linear programming can be used to establish the existence of a non-negative solution w_t to this system such that $\sum_{i=1}^{k} w_t^i = 1$.

Lemma 1 (No-regret) *For all i^* and j^* the regret grows as the squareroot of T. In particular,* $R_T^{i^*\to j^*} \le \sqrt{2kT}$.

Proof: Let $G_\epsilon(x) \equiv \dfrac{\epsilon x^2}{2} I_{x>0}$. Since $x \le \frac{1}{2\epsilon} + G_\epsilon(x)$ we see that

$$R_T^{i^*\to j^*} \le \frac{1}{2\epsilon} + G_\epsilon(S_T^{i^*j^*}) \le \frac{1}{2\epsilon} + \sum_{ij} G_\epsilon(S_T^{ij})$$

Now $G'_\epsilon(x) = \epsilon x I_{x>0}$ and so

$$\begin{aligned}
\sum_{ij} (S_t^{ij} - S_{t-1}^{ij}) G'_\epsilon(S_{t-1}^{ij}) &= \sum_{ij} w_t^i (L_t^i - L_t^j)(\epsilon S_{t-1}^{ij} I_{S_{t-1}^{ij}>0}) \\
&= \epsilon \sum_i L_t^i \underbrace{\Big(w_t^i \sum_j R_{t-1}^{i\to j} - \sum_j w_t^j R_{t-1}^{j\to i} \Big)}_{=0 \text{ by flow conservation}} \\
&= 0
\end{aligned}$$

Expanding $G_\epsilon(S_t^{ij})$ as a two term Taylor series around S_{t-1}^{ij} shows

$$\sum_{ij} G_\epsilon(S_t^{ij}) \;\le\; \sum_{ij} G_\epsilon(S_{t-1}^{ij}) + \sum_{ij} (S_t^{ij} - S_{t-1}^{ij}) G'_\epsilon(S_{t-1}^{ij}) + \sum_{ij} (w_t^i)^2 (L_t^i - L_t^j)^2 \epsilon$$

$$
\begin{aligned}
&\leq \sum_{ij} G_\epsilon(S_{t-1}^{ij}) + k\epsilon \sum_i (w_t^i)^2 \\
&\leq \sum_{ij} G_\epsilon(S_{t-1}^{ij}) + k\epsilon.
\end{aligned}
$$

Computing the recursive sum we see that $\sum_{ij} G_\epsilon(S_T^{ij}) \leq Tk\epsilon$ and so $R_T^{i^*\to j^*} \leq \frac{1}{2\epsilon} + Tk\epsilon$. Picking $\epsilon = 1/\sqrt{2kT}$ shows $R_T^{i^*\to j^*} \leq \sqrt{2Tk}$. □

We will now show that for a suitable loss function, a randomized forecast that has no regret must also be calibrated.

- First, our forecasting scheme will choose in each round a probability vector from the set $\{p^i | i = 0, 1, \ldots, k\}$ which is chosen so that any probability distribution over S(2) (the opponents strategies) is within δ of one of these points.
- We denote the move made by player in 2 by the vector $X_t = [X_{t,1}, X_{t,2}, X_{t,3}, \ldots]$ where $X_{t,j} = 1$ if strategy $j \in S(2)$ was chosen and zero otherwise. Notice that X_t will be a 0-1 vector with exactly on non-zero component.
- Next, the loss incurred in round t from forecasting p^i will be $L_t^i = |X_t - p^i|^2 = \sum_{j\in S(2)} |X_{t,j} - p_j^i|^2$.
- The probability of forecasting p^i at time t will be w_t^i.

We would like to choose the w_t's so that L-2 calibration $C_2(t)$ goes to zero in probability as t gets large, where

$$C_2(t) = \sum_p (\rho(p, j, t) - p_j)^2 \frac{N(p, t)}{t}$$

The expected value of $C_2(t)$ is given by:

$$E(C_2(t)) = \sum_{s=1}^{t} \sum_{i=1}^{k} \sum_{j\in S(2)} w_t^i (\rho_t(p^i, j, s) - p_j^i)^2 / s.$$

Simple algebra yields

$$\sum_i \max_j R_t^{i\to j}/t \le E(C_2(t)) \le \delta + \sum_i \max_j R_t^{i\to j}/t$$

If the probabilities w_t's are chosen to satisfy the flow conservation equations displayed earlier, we deduce that

$$E(C_2(t)) \le \delta + O(\frac{k}{\sqrt{t}}).$$

Thus if we let k grow slowly and δ go slowly to zero, we see that $C_2(t) \to 0$ in expectation which implies $C_2(t) \to 0$ in probability by Jensen's inequality. The L-1 calibration definition of equation (1) follows from the fact that it is smaller than the square root of the L-2 calibration. Thus we have proved Theorem 3.

REFERENCES

Aumann, R. J., "Subjectivity and Correlation in Randomized Strategies", *Journal of Mathematical Economics*, **1**, 67-96, 1974.

Aumann, R. J., "Correlated Equilibrium as an Expression of Bayes Rationality", *Econometrica*, **55**, #1, 1-18, 1987.

Blackwell, D., "A vector valued analog of the mini-max theorem," *Pacific J. Math.* **6**, 1-8, 1956.

Dawid, A. P., "The Well Calibrated Bayesian", *Journal of the American Statistical Association,* **77**, #379, 605-613, 1982.

Foster, D. P., and R. Vohra "Asymptotic Calibration", unpublished manuscript, 1991.

Foster, D. P., and H. P. Young "Stochastic Evolutionary Game Dynamics," *Theoretical Population Biology,* **38**, 219-232, 1990.

Fudenberg, D. and D. Kreps, "Experimentation, Learning, and Equilibrium in Extensive Form Games", unpublished notes, 1991.

Fudenberg, D. and D. Levine, "An easier way to Calibrate", unpublished manuscript, 1995.

Fudenberg, D. and D. Levine, "Conditional Universal Consistency", unpublished manuscript, 1996.

Hart, S. and Mas-Collel, A., "A Simple Adaptive Procedure Leading to Correlated Equilibrium," unpublished manuscript, 1996.

Kalai, E. and E. Lehrer, "Subjective Games and Equilibria", unpublished manuscript, 1994.

Kalai, E., E. Lehrer and R. Smorodinsky, "Calibrated Forecasting and Merging", manuscript, 1996.

Kreps, D. "Game Theory and Economic Modeling," Oxford University Press, Oxford, 1991a.

Kreps, D., Seminar on Learning in Games, University of Chicago, 1991b.

Milgrom, P. and J. Roberts "Adaptive and Sophisticated Learning in Normal Form Games", *Games and Economic Behavior*, 3, 82-100, 1991.

Monderer, D. and L. Shapley "Fictitous Play Property for games with Identical Interests," working paper, Department of Economics, UCLA, 1993.

Miyasawa, K., "On the Convergence of the Learning Process in a 2×2 Non-zero- sum Two-person Game," Economic Research Program, Princeton University, Research Memorandum #33 (1961).

Nau, R. F. and K. McCardle, "Coherent Behavior in Non-cooperative Games," *Journal of Economic Theory*, **50**, #2, 1990.

Oakes, D. "Self-Calibrating Priors Do Not Exist", *Journal of the American Statistical Association*, **80**, 339, 1985.

Robinson, J. "An Iterative Method of Solving a Game" *Annals of Math-*

ematics, **54**, 296-301, 1951.

Samuelson, L. and J. Zhang, "Evolutionary Stability in Asymmetric Games", *Journal of Economic Theory*, **57**, 363-391, 1992.

Shapley, L. "Some Topics in Two-Person Games," *Advances in Game Theory*, Princeton University Press, Princeton 1964.

Skyrms, B. *The Dynamics of Rational Deliberation*, Harvard University Press, 1990.