GAMES AND ECONOMIC BEHAVIOR **22,** 274–298 (1998)
ARTICLE NO. GA970589

Learning to Play Limited Forecast Equilibria

Philippe Jéhiel*

C.E.R.A.S.,† Ecole Nationale des Ponts et Chaussées, 28 rue des Saints-Pères, 75007 Paris, France and University College London, United Kingdom

Received November 14, 1994

This paper provides a learning justification for limited forecast equilibria, i.e., strategy profiles such that (1) players choose their actions in order to maximize the discounted average payoff over their horizon of foresight as given by their forecasts and (2) forecasts are correct on and off the equilibrium path. The limited forecast equilibria appear to be the stochastically stable outcomes of a simple learning process involving (vanishing) trembles. *Journal of Economic Literature* Classification Numbers: C72, D83. © 1998 Academic Press

1. INTRODUCTION

Several approaches to bounded rationality in repeated games have been considered so far. A first approach is concerned with the complexity of the strategies used by the players (Neyman, 1985; Rubinstein, 1986; Kalai and Stanford, 1988), and some authors suggest including complexity concerns in the objective of the players (Rubinstein, 1986; Abreu and Rubinstein, 1988). Another approach restricts the attention to strategies with bounded recall (for example, Lehrer, 1988), or combines complexity ideas with bounded recall ideas (Kalai and Stanford, 1988). Finally, some of the learning (or the evolutionary game) literature assumes that the players are myopic even though they act in a long-run environment (for example, Jordan, 1991).

Jéhiel (1995) considers an alternative approach to bounded rationality taking the view that when the horizon is too long individuals are unlikely to be able to correctly forecast the entire future. Individuals are assumed to form predictions about what will happen in a *limited* horizon future.

* I thank Tilman Börgers, Olivier Compte, Eddie Dekel, Ehud Kalai, Bart Lipman, Sylvain Sorin, Jörgen Weibull, seminar participants at the CEPR Summer meeting, Gerzensee 1995, and at a seminar in Paris (Ecole Normale Supérieure), as well as an associate editor and two anonymous referees for helpful comments on earlier drafts of this paper.

† Unité de Recherche CNRS (URA 2036). E-mail: jehiel@enpc.fr.

0899-8256/98 $25.00

They subsequently make their decisions on the basis of their limited horizon forecast. Specifically, Jéhiel (1995) considers two-player repeated alternate-move games with arbitrary finite action spaces A_i, $i = 1, 2$. Each player $i = 1, 2$ repeatedly makes his choice of current action (in A_i) on the basis of his limited n_i length-forecasts.[1] The limited forecast equilibrium is referred to as (n_1, n_2)-solution and defined as a strategy profile such that (1) players choose their actions so as to maximize the discounted average payoff over their horizon of foresight and (2) the limited horizon forecasts formed by the players are correct on and off the equilibrium path. It can be shown that there always exists at least one (n_1, n_2)-solution and that the period t limited horizon forecasts of any (n_1, n_2)-solution repeat cyclically as the time period t varies. The length of a cycle induced by any (n_1, n_2)-solution can be bounded by K, where K depends on the lengths of foresight n_i and the cardinality of the action spaces A_i only (see Jéhiel, 1995).

The objective of this paper is to provide a learning justification for the correctness of equilibrium forecasts on and off the equilibrium path. We follow Kalai and Lehrer (1993a) in that the learning process takes place within the play of the game. Initially each player i has a belief over several possible forecasting rules, which are sequences of n_i-length forecasts one for each period where this player must move. At each period the player who must move either (1) selects an action based on his belief so as to maximize the discounted average payoff over the next n_i periods or (2) trembles with a small probability, and may choose any action with positive probability (see Selten, 1975). Player i subsequently observes the played actions in the past periods, gathers them into n_i-length streams of actions, and compares the latter with the predictions associated with each of the possible forecasting rules, which in turn allows him to update his belief. Specifically, when the prediction of a forecasting rule does not coincide with the observation, then some tremble must have occurred to explain the observation with that forecasting rule. Such a forecasting rule becomes a little less plausible relative to those forecasting rules whose prediction fits with the observation. Besides, we assume that each player restricts himself to a limited (though arbitrarily large) number of plausibility levels (to be defined below). Also, when the player cannot discriminate which forecasting rule is the most plausible one, we assume that with positive probability he may change his state of belief resulting in a possibly new most plausible forecasting rule. Finally, we assume that the supports of initial beliefs of the players are finite and contain all cyclical forecasting rules with a length of cycle less than or equal to K, which ensures that the forecasting rule of

[1] Rational and myopic behavior correspond to an infinite and zero length of foresight, respectively.

at least one (n_1, n_2)-solution is contained in the initial support of each player.[2]

The main result of this paper is that, as the probability of trembling converges to zero, the players eventually play almost surely as in a (n_1, n_2)-solution. In other words, the stochastically stable outcomes of the process defined by the above learning story coincide with the (n_1, n_2)-solutions.

To the best of our knowledge, this is the first attempt to justify a bounded rationality solution concept as a result of a two-player-learning process where the learning takes place during the play of the game. This should be contrasted, for example, with Binmore and Samuelson (1992), who apply evolutionary ideas (population learning) to the finite automaton framework (and obtain results that differ from the solution concept proposed by Rubinstein, 1986, and Abreu and Rubinstein, 1988).

Technically, the analysis borrows from the pioneering work of Foster and Young (1990) which was further developed and applied by Kandori *et al*. (1993), Young (1993a, b), Fudenberg and Harris (1992), Nöldeke and Samuelson (1993), Kandori and Rob (1995), and others (see Kandori, 1996, for a survey). Those works were primarily applied to evolutionary contexts, and the noisy character of the process was interpreted as a probability of mutation rather than a probability of tremble. It turns out that similar techniques can be applied to our framework too. Intuitively, the stochastically stable outcomes correspond to the absorbing sets of the process without trembles which are hardest to destabilize, i.e., from which it is hardest to get out. In our framework, the absorbing sets of the process *without trembles* correspond to the self-confirming (n_1, n_2)-solutions, i.e., strategy profiles such that the associated n_i-length forecasts are correct only on the equilibrium path (see Fudenberg and Levine, 1993, and Kalai and Lehrer, 1993b, for a similar concept in a framework with perfect rationality). We next observe that destabilizing a (n_1, n_2)-solution requires several *non-isolated* trembles, whereas *isolated* trembles are enough to destabilize a self-confirming (n_1, n_2)-solution (that is not a (n_1, n_2), solution). Since the former type of events is much more likely than the latter, it follows that, in the limit as the probability of trembling converges to zero, a (n_1, n_2)-solution is eventually played with probability 1.

In Section 2 the model is described. The solution concept is defined in Section 3. Sections 4 and 5, respectively, present and analyze the learning

[2] It should be noted that in general there will be forecasting rules in the support of the players' belief such that the forecast at some period is inconsistent with the forecast at a later period. That is, the consistency attached to the infinite horizon of the game is not required at the individual level.

process. A discussion follows in Section 6. Section 7 concludes. All proofs are gathered in the Appendix.

2. THE MODEL

We consider discounted *repeated alternate-move games* with two players $i = 1, 2$. Player i chooses actions a_i from a finite action space A_i. Players act in discrete time, and the horizon is infinite. Periods are indexed by t $(t = 1, 2, 3, \dots)$. At time t, player i's single period payoff is a function of the current actions a_i^t of the two players $i = 1, 2$, but not of time: $u_i = u_i(a_1^t, a_2^t)$. Players move sequentially and player 1 moves first. At each odd period $t = 2k - 1$ $(t = 1, 3, 5, \dots)$, player 1 chooses an action that remains the same for the two periods t and $t + 1$: $a_1^{2k} = a_1^{2k-1}$ for all k. Similarly, player 2 moves at each even period $t = 2k$ $(t = 2, 4, 6, \dots)$ and $a_2^{2k+1} = a_2^{2k}$. Each player $i = 1, 2$ discounts the future. The discount factor of player i is denoted by δ_i.

A stream of action profiles $\{q_i^t\}_{t=1}^{\infty} = \{q_1^{2k-1}, q_2^{2k}\}_{k=1}^{\infty}$, where $q_1^{2k-1} \in A_1$ and $q_2^{2k} \in A_2$ is referred to as a path and is denoted by Q. Since players may only change actions every other period, a move at period t affects payoffs both at periods t and $t + 1$. In path Q, each action q_2^{2k} (resp. q_1^{2k+1}) of player 2 (resp. 1) is thus combined both with the previous action q_1^{2k-1} (resp. q_2^{2k}) and the next action q_1^{2k+1} (resp. q_2^{2k+2}) of player 1 (resp. 2): At periods $2k$ and $2k + 1$, the *current* payoffs to player i induced by path Q are $u_i(q_1^{2k-1}, q_2^{2k})$ and $u_i(q_1^{2k+1}, q_2^{2k})$, respectively.[3] We first introduce some preliminary and standard notation.

Notation. (1) Let R_n denote an arbitrary n-length stream of alternate actions. $v_i(R_n)$ denotes the discounted sum of the per period payoffs to player i induced by R_n where each action of R_n is combined both with the previous (except for the first one) and the next (except for the last one) action of R_n. For example the 4-length stream $R_4 = (a_1, a_2, b_1, b_2)$, where $a_i, b_i \in A_i$ induces: $v_i(R_4) = v_i(a_1 a_2 b_1 b_2) = u_i(a_1, a_2) + \delta_i u_i(b_1, a_2) + (\delta_i)^2 u_i(b_1, b_2)$.

(2) $[Q]_n$ denotes the truncation of path, $Q = \{q_i^t\}_{t=1}^{\infty}$, to the *first n* actions, i.e., $[Q]_n = \{q_i^t\}_{t=1}^{n}$, and $[Q]_{t'}^{t''} = \{q_i^t\}_{t=t'}^{t=t''}$ is the associated stream of actions from period t' to period t''.

(3) (q, q') denotes the concatenation of $q = \{q_i^t\}_{t=t'}^{t''}$ with $q' = \{q_i^t\}_{t=t''+1}^{t'''}$: $(q, q') = \{q_i^t\}_{t=t'}^{t'''}$.

[3] Single period payoffs start at period 2.

3. THE LIMITED FORECAST EQUILIBRIUM

Players are assumed to make limited predictions about the forthcoming moves after their own move. Player i only considers the forthcoming n_i moves after his own move, and subsequently makes his choice of current action on the sole basis of his predictions. Jéhiel (1995) introduces a solution concept along this line where the predictions made by the players may depend on the past N actions together with the currently played action and the time period. The dependence on the past N actions can be shown to play no role as far as the set of solutions is concerned (Jéhiel, 1995), and the analysis of the learning process (see below) could trivially be extended to that case. For notational convenience, we will therefore assume that the limited predictions formed by player i depend only on the action to be currently chosen by this player and the time period. We now introduce some definitions and notation.

Definitions and Notation

(1) A n_i-length (pure) prediction for player i is a stream of alternate actions of length n_i starting with an action in A_j $(j \neq i)$. The set of n_i-length predictions is denoted by P_{n_i}, where $P_{n_i} = (A_j \times A_i)^{n_i/2}$ if n_i is even and $P_{n_i} = A_j \times (A_i \times A_j)^{(n_i-1)/2}$ if n_i is odd.

(2) A n_i-length forecast for player i at a period t where this player must move is denoted by f_i^t. It maps the set of actions A_i to be currently chosen into the set of predictions P_{n_i}. Formally, f_i^t: $A_i \to P_{n_i}$ where $f_i^t(a_i) \in P_{n_i}$ is the prediction about the forthcoming n_i actions made by player i at period t if he currently chooses a_i.

(3) $f_i = \{f_i^t\}_t$ is a forecasting rule. It is a sequence of forecasts f_i^t one for each period t where player i must move. The set of f_i is denoted $\mathscr{F}_i$. A forecasting rule profile $(f_1, f_2) \in \mathscr{F}_1 \times \mathscr{F}_2$ is denoted by f, and the set of f is denoted $\mathscr{F}$.

(4) A pure strategy for player i is denoted by σ_i. It is a sequence of functions σ_i^t, one for each period t where player i must move. The function at period t, σ_i^t, is the behavior strategy of player i at that period. It determines player i's action at period t as a function of the last action played by j. Formally, σ_i^t: $A_j \to A_i$.[4] The set of player i's strategies is denoted by Σ_i. A strategy profile (σ_1, σ_2) is denoted by σ, and the set of strategy protiles $\Sigma_1 \times \Sigma_2$ is denoted Σ.

[4] σ_i^t may only depend on the last action because forecasts are assumed to be history-independent. Observe that the period 1 behavior strategy does not depend on the previous action since there is no such action, and therefore $\sigma_1^1 \in A_1$.

Any strategy profile $\sigma \in \Sigma$ generates a path denoted by $Q(\sigma) = \{q_i^t(\sigma)\}_t$, $i = 1$ (resp. 2) if t is odd (resp. even). Let $\mathscr{H}^t$ denote the set of histories of alternate actions of length t. Let h be an arbitrary history of length $t-1$, i.e., $h \in \mathscr{H}^{t-1}$. The strategy profile and the path induced by σ on the subgame following h are denoted by $\sigma|_h$ and $Q(\sigma|_h)$, respectively. Given $h \in \mathscr{H}^{t-1}$ and the action $a_i \in A_i$ at period t, the continuation path induced by σ after (h, a_i) is thus $Q(\sigma|_{ha_i})$.

The limited forecast solution concept requires two conditions. First, players use the discounted average per period payoff over the length of foresight as their criterion to select current actions:

DEFINITION 1. A strategy $\sigma_i \in \Sigma_i$ is *justified* by the forecasting rule $f_i = \{f_i^t\}_t \in \mathscr{F}_i$ if

$$\forall t, \forall a_j \in A_j \qquad \sigma_i^t(a_j) \in \operatorname*{Arg\,max}_{a_i} \nu_i\left(a_j a_i f_i^t(a_i)\right),$$

where a_j stands for the period $t-1$ action of player j, $j \neq i$.

Throughout the paper we will consider generic values of the payoffs in the sense that for $i = 1, 2$ there is no $a_j \in A_j$, $a_i \in A_i$, $a_i' \neq a_i \in A_i$, $p_i \in P_{n_i}$, $p_i' \neq p_i \in P_{n_i}$ such that $\nu_i(a_j a_i p_i) = \nu_i(a_j a_i' p_i')$. Thus, in Definition 1, $\operatorname{Arg\,max}_{a_i} \nu_i(a_j a_i f_i^t(a_i))$ is a singleton, and $\sigma_i^t(a_j) = \operatorname{Arg\,max}_{a_i} \nu_i(a_j a_i f_i^t(a_i))$.

The second requirement is that players' equilibrium forecasts are related to equilibrium strategies by a consistency relationship, where consistency means that forecasts are correct *on* and *off* the equilibrium path, i.e., for every (h, a_i) the period t forecast if player i chooses a_i, $f_i^t(a_i)$, coincides with the truncation to the first n_i actions of the continuation path induced by σ, $[Q(\sigma|_{ha_i})]_{n_i}$.

DEFINITION 2. $f_i = \{f_i^t\}_t \in \mathscr{F}_i$ is *consistent* with $\sigma \in \Sigma$ if for every period t where player i must move: $\forall a_i \in A_i$, $\forall h \in \mathscr{H}^{t-1}$, $f_i^t(a_i) = [Q(\sigma|_{ha_i})]_{n_i}$.

To summarize, a (n_1, n_2)-solution is a strategy profile that can be *justified* by *consistent* forecasting rules for players 1 and 2:

DEFINITION 3 (The Solution Concept). A strategy profile $\sigma = (\sigma_1, \sigma_2) \in \Sigma$ is a (n_1, n_2)-solution if and only if there exists a forecasting rule profile $f = (f_1, f_2) \in \mathscr{F}$ such that, for $i = 1, 2$,

(1) σ_i is *justified* by f_i

(2) f_i is *consistent* with σ.

The forecasting rule profile f appearing in Definition 3 is uniquely defined given the strategy profile σ; we say that f is *associated* with σ. The properties of (n_1, n_2)-solutions are analyzed in Jéhiel (1995). In particular, it is shown that (n_1, n_2)-solutions always exist and can be constructed backwards. Moreover, the forecasting rule f_i associated with a (n_1, n_2)-solution is cyclical,[5] e.g., $\exists k$ s.t. $\forall t$, $f_i^t(\cdot) = f_i^{t+k}(\cdot)$. The minimal k such that the latter property holds for the equilibrium forecasting rules of both players is referred to as the length of the cycle induced by the (n_1, n_2)-solution. It can be shown that the length of the cycle of any (n_1, n_2)-solution is no greater than an upper bound $K(n_1, n_2)$ which depends only on the lengths of foresight n_i (and the cardinality of the action spaces $|A_i|$).[6]

The consistency requirement introduced in Definition 2 assumes that forecasts are correct on and off the equilibrium path. For the analysis of the learning process it will be convenient to introduce an alternative (weaker) notion of consistency (termed *subjective* consistency) for which forecasts are correct *on* the equilibrium path but not necessarily *off* the equilibrium path (see Fudenberg and Levine, 1993, and Kalai and Lehrer, 1993b). The weaker notion of consistency leads in turn to a weaker solution concept that we call *self-confirming* (n_1, n_2)-solution. Formally,

DEFINITION 4 (Subjective Consistency). $f_i = \{f_i^t\}_t \in \mathcal{F}_i$ is *subjectively* consistent with $\sigma \in \Sigma$ if, for every period t where player i must move, $(h, a_i) = [Q(\sigma)]_t \Rightarrow f_i^t(a_i) = [Q(\sigma|ha_i)]_{n_i}$.

DEFINITION 5 (Self-Confirming Limited Forecast Equilibrium). A strategy profile $\sigma = (\sigma_1, \sigma_2) \in \Sigma$ is a *self-confirming* (n_1, n_2)-solution if and only if there exists a forecasting rule profile $f = (f_1, f_2) \in \mathcal{F}$ such that, for $i = 1, 2$,

(1) σ_i is justified by f_i

(2) f_i is *subjectively* consistent with σ.

Finally, for the following analysis it will be convenient to introduce the following notation: $Q^t(f; a_i^t)$ will denote the stream of actions from period t on induced by the period t action a_i^t and the forecasting rule profile $f = (f_1, f_2)$, where each player i selects in time the action that maximizes the discounted average payoff over the forthcoming n_i periods as given by his forecasting rule f_i.[7] $Q^*(f)$ will denote the set of all paths $Q^1(f; a_1^1)$ generated by the forecasting rule profile $f = (f_1, f_2)$ with arbitrary first period actions a_1^1.

[5] Jéhiel (1995) considers the case where there is no discounting. However, the mentioned properties trivially extend to the case with discounting.

[6] Specifically, $K(n_1, n_2) = (\text{Max}(|A_1|, |A_2|)^{\text{Max}(n_1, n_2)+1}$; see Jéhiel (1995).

[7] This path is uniquely defined because of the genericity assumption.

4. LEARNING AND LIMITED FORECASTING

The objective of this paper is to provide a justification for why forecasts should be correct as a result of learning. The basic view is that each player i restricts his endeavor to trying to understand what the forthcoming n_i moves will be as a function of his current action. To this end, he forms a belief over forecasting rules where forecasting rules specify how to make n_i-length predictions in the dynamic environment of the game. At each period the player who must move either (1) selects an action based on his belief or (2) trembles, i.e., makes a mistake. The mistakes occur with a small probability, and the player may then choose any action with positive probability. When he does not tremble, player i selects an action which maximizes the discounted average payoff over the next n_i periods according to his belief. The selection is typically based on that (or those) forecasting rule which is currently the most plausible one. Player i subsequently observes the played actions in the past periods. As soon as a new n_i-length prediction can be compared with a realized stream of actions, player i asks himself, for each possible forecasting rule f_i he may consider, whether the prediction associated with f_i is compatible with the observation or whether some mistake is required to explain the observation with f_i.[8] A forecasting rule whose prediction does not fit with the observation becomes a little less plausible relative to a forecasting rule whose prediction fits with the observation. Also, each player i is assumed to restrict himself to a finite number of plausibility levels (to be defined below), and when there are several forecasting rules that are candidates for being the most plausible one, the player may change his state of belief with positive probability resulting in a possibly new most plausible forecasting rule.

The main result of the paper is that provided the support of the players' initial belief is finite and contains all forecasting rules which have a cycle of length no greater than K, where $K > K(n_1, n_2)$ (so that the forecasting rules of (n_1, n_2)-solutions belong to the supports of initial beliefs, see the end of Section 3), we are sure that, as the probability of making a mistake goes to zero, a (n_1, n_2)-solution is eventually played with probability 1.

Before we describe the learning process, we wish to point out that the above learning story does *not* require a great sophistication on the part of the players. First, even though the set of all forecasting rules is quite large, the support of player i's initial belief is *not* required to contain infinitely many of these.[9] Specifically, it is required to be finite and contain all cyclical forecasting rules of length no greater than K. The finiteness of the

[8] The underlying idea is that when player i looks at the previously played actions (including his own actions) he does not know whether those result from trembling or not.

[9] This should be contrasted with Kalai and Lehrer (1993a), who cannot *a priori* assume the initial supports of the players to be finite; see also Nachbar (1997).

support of player i's belief seems a desirable assumption given that a boundedly rational player may not be able to keep track of the plausibility of infinitely many forecasting rules. Also, because cyclical forecasting rules of length no greater than K have a simple structure, they are more likely to be considered in the support of belief of the players.

Second, it should be noted that some forecasting rules (including cyclical ones) are such that the n_i-length forecast of one period is not consistent with that of a future period.[10] Each player i is allowed to assign positive weight to such forecasting rules which are dynamically inconsistent, and therefore we impose no restriction on the dynamic consistency of the forecasting rules to be considered by the players. It turns out, however, that inconsistent forecasting rules will eventually appear to be less plausible as a result of learning because they are less compatible with the observations. Further comments about the learning process will be presented in Section 6.

4.1. *The Learning Process*

The Mistakes

At each period where he must move, player i may tremble with probability ε (where ε should be thought of as being small). The trembles at each period are independent from each other and stationary (i.e., independent of the history of plays). When he trembles, player i may choose any action $a_i \in A_i$ with a positive probability (assumed to be independent of ε). For example, each action $a_i \in A_i$ may then be played with the same probability $1/|A_i|$. It should be pointed out, however, that the choice of a specific distribution is immaterial for the asymptotic results to be described below as long as every action is played with a strictly positive probability.

State of Belief and Forecasting Rules

Each player i has a belief over a finite support of forecasting rules denoted by F_i, where F_i is assumed to include the set of all cyclical forecasting rules with a cycle length no greater than K, $K > K(n_1, n_2)$ (see above). For simplicity, we will present the argument for the case where the support F_i consists *only* of cyclical forecasting rules, i.e., $\forall f_i \in F_i$, $\exists k$, s.t. $\forall t$, $f_i^{t+k}(\cdot) = f_i^t(\cdot)$), but we do not require the cycle of $f_i \in F_i$ (i.e., k) to be necessarily smaller than K. (The analysis could easily be extended to the case where F_i contains also forecasting rules that are cyclical only after some time period.) The state of belief of player i is meant to represent the

[10] For example, if $n_i = 3$, $f_i^t(a_i) = a_j^{t+1}a_i^{t+2}a_j^{t+3}$ and $f_i^{t+2}(a_i^{t+2}) = a_j'^{t+3}a_i^{t+4}a_j^{t+5}$ with $a_j^{t+3} \neq a_j'^{t+3}$.

plausibility of every forecasting rule $f_i \in F_i$ based on past observations. We assume that player i restricts himself to k_i levels of plausibility where k_i is assumed to be sufficiently large, i.e., no smaller than $n_i + 1$. Forecasting rules which belong to level 1 are the most plausible ones, those in level 2 are the second most plausible ones, and so on till level k_i, which contains the least plausible forecasting rules. A state of belief for player i is denoted by s_i. It maps the support of forecasting rules F_i into the set $\{1, \ldots, k_i\}$. Formally, s_i: $F_i \to \{1, \ldots, k_i\}$, where $s_i(f_i)$ is the plausibility level of forecasting rule f_i. We denote by $S_i(k) = \{f_i \in F_i / s_i(f_i) = k\}$, and by $\mathscr{S}_i$ the set of all states of beliefs s_i such that $S_i(1)$ has cardinality 1, i.e., for which there is one and only one most plausible forecasting rule.

Behavior and State of Belief

When player i does not tremble at period t his choice of action is determined by his current (or most recently formed, see below the timing of updating) state of belief s_i and the action a_j^{t-1} played by player j at period $t - 1$. The state of belief s_i to be considered by player i will always be such that $S_i(1)$ has cardinality 1 (see the discussion section below), and therefore $s_i \in \mathscr{S}_i$. We let f_i^* denote the most plausible forecasting rule according to s_i, i.e., the only forecasting rule $f_i \in F_i$ such that $s_i(f_i) = 1$. When he does not tremble, player i selects an action a_i^t so as to maximize the discounted average payoff over his horizon of forecast as given by his currently most plausible forecasting rule f_i^*. That is, he selects

$$a_i^t = \operatorname*{Arg\,max}_{a_i \in A_i} \nu_i\left(a_j^{t-1} a_i f_i^{*t}(a_i)\right)$$

(which is uniquely defined given the genericity assumption).

Updating the State of Belief

A the end of period $t + n_i$, where period t is a period where player i has moved, player i may check his period t forecast for every forecasting rule $f_i \in F_i$. A new state of belief can then be formed. Let s_i^{t-2} denote the state of belief of player i that prevails immediately before period $t + n_i$ (it has thus been formed in period $t + n_i - 2$). Let a_i be the action played at period t, and let $[h]_{t+1}^{t+n_i}$ be the stream of actions played from period $t + 1$ to period $t + n_i$. Player i's new state of belief is denoted by s_i^t. It is updated by comparing the prediction $f_i^t(a_i)$ of every forecasting rule $f_i \in F_i$ with the realized n_i-length stream $[h]_{t+1}^{t+n_i}$. Specifically, the new state of belief s_i^t is derived from the function $\bar{s}_i^t$ defined by

$$\bar{s}_i^t(f_i) = \begin{cases} s_i^{t-2}(f_i) & \text{if } f_i^t(a_i) = [h]_{t+1}^{t+n_i} \\ s_i^{t-2}(f_i) + 1 & \text{if } f_i^t(a_i) \neq [h]_{t+1}^{t+n_i} \end{cases}$$

(it adds one increment if the prediction is incorrect). The state of belief s_i^t is determined on the basis of $\bar{s}_i^t$, but s_i^t is required to be an element of $\mathscr{S}_i$ so that some transformation is required. We let f_i^* denote the most plausible forecasting rule according to $s_i^{t-2} \in \mathscr{S}_i$, i.e., $s_i^{t-2}(f_i^*) = 1$. We note that

$$f_i^* \in \underset{f_i \in F_i}{\text{Arg Min}}\ \bar{s}_i^t(f_i).$$

Case 1 (f_i^* *Is the Only Forecasting Rule Minimizing* $\bar{s}_i^t(\cdot)$). Then f_i^* remains the most plausible forecasting rule, and $s_i^t(f_i) = \text{Min}(1 + \bar{s}_i^t(f_i) - \bar{s}_i^t(f_i^*), k_i)$ for all $f_i \in F_i$. In other words, when f_i^* yields a *correct* prediction, the plausibility level of $f_i \in F_i$ increases by one increment if the maximum level k has not been reached yet and the prediction of f_i is incorrect; it remains the same otherwise. When f_i^* yields an *incorrect* prediction, the plausibility level of $f_i \in F_i$ decreases by one increment if the prediction of f_i is correct, and remains the same otherwise.

Case 2 (*There Are Several Forecasting Rules Minimizing* $\bar{s}_i^t(\cdot)$). Then player i's state of belief switches to some new state of belief $s_i^t \in \mathscr{S}_i$ according to some distribution assumed to assign positive weight to every state of belief $s_i \in \mathscr{S}_i$. Note that the distribution from which s_i^t is drawn may in general depend on the function $\bar{s}_i^t$. For example, with probability $1 - \varepsilon'$ the new most plausible forecasting rule may be one of the forecasting rules

$$f_i^{**} \in \underset{f_i \in F_i}{\text{Arg Min}}\ \bar{s}_i^t(f_i) \qquad \big(\text{i.e., } s_i^t(f_i^{**}) = 1\big),$$

the plausibility level of other forecasting rules being updated accordingly, i.e., $s_i^t(f_i) = 2$ if $f_i \in \text{Arg Min}\ \bar{s}_i^t(\cdot)$ and $f_i \neq f_i^{**}$; $s_i^t(f_i) = \text{Min}(1 + \bar{s}_i^t(f_i) - \bar{s}_i^t(f_i^*), k_i)$ otherwise. With probability ε' any state of belief $s_i^t \in \mathscr{S}_i$ is equally likely to arise.

Initialization

The above elements of the learning process implicitly define a stochastic process. A global (current) state in this stochastic process is denoted by g; it consists of (1) a pair of (current) state of belief s_i for each player $i = 1, 2$ and (2) a stream of $\text{Max}(n_1, n_2) + 2$ (or more) alternate actions (standing for the last played actions). The set of states g is finite. It is denoted by $\mathbb{G}$. The transition from state to state occurs every other period. The process is initialized by considering some arbitrary initial global state. (The particular choice of an initial state plays no role in the analysis.)

5. LEARNING TO PLAY (n_1, n_2)-SOLUTIONS

Two classes of states in $\mathbb{G}$ will play a special role in the analysis. The class of (n_1, n_2)-solution states, and the class of self-confirming (n_1, n_2)-solution states which are defined by:

DEFINITION 6. A state $g \in \mathbb{G}$ is a (n_1, n_2)-solution *state* (resp. self-confirming (n_1, n_2)-solution *state*) if there exists a forecasting rule profile $f = (f_1, f_2)$ associated with a (n_1, n_2)-solution (resp. self-confirming (n_1, n_2)-solution) $\sigma \in \Sigma$ together with a path $Q \in Q^*(f)$ generated by $f = (f_1, f_2)$ such that (1) f_i is player i's most plausible forecasting rule according to player i's state of belief s_i (i.e., $s_i(f_i) = 1$), (2) player i's forecasting rules $f_i \in F_i$ which yield some incorrect predictions along the path Q (those repeat cyclically because everything is cyclical) are among the least plausible forecasting rules for player i (i.e., they belong to $S_i(k_i)$), and (3) the stream of $\text{Max}(n_1, n_2) + 2$ actions (in g) coincides with $[Q]_{t+1}^{t+\text{Max}(n_1, n_2)+2}$ for some t (it corresponds to $\text{Max}(n_1, n_2) + 2$ consecutive actions in Q).

In the following analysis, it will be convenient to gather (self-confirming) (n_1, n_2)-solution states which have the same states of belief for players 1 and 2 but which may differ in their streams of $\text{Max}(n_1, n_2) + 2$ actions (due to the position of the cycle in the path Q generated by the (self-confirming) (n_1, n_2)-solution). Such sets will be referred to as clusters:

DEFINITION 7. The set of (n_1, n_2)-solution (resp. self-confirming (n_1, n_2)-solution) states which correspond to the same states of belief for each player $i = 1, 2$ and which may differ only in their stream of $\text{Max}(n_1, n_2) + 2$ actions is referred to as a (n_1, n_2)-solution (resp. self-confirming (n_1, n_2)-solution) *cluster*. The *set* of (n_1, n_2)-solution clusters is denoted $\mathbb{E}$, and the *set* of self-confirming (n_1, n_2)-solution clusters which are not in $\mathbb{E}$ is denoted $\mathbb{S}$.

5.1. *Absorbing Sets without Trembles*

We first study the learning dynamics in the absence of trembles (i.e., $\varepsilon = 0$). The learning process defines a Markov process, and we are interested in the stationary distributions of this Markov process. A set of states is absorbing if it is a minimal set of states with the property that the Markov process can lead into this set but not out of it. An absorbing set may *a priori* contain only a single state in which case it is a stationary state of the Markov process or it may contain more than one state in which case the Markov process cycles between states in the absorbing set. It is readily verified that starting from a state in a self-confirming (n_1, n_2)-solution cluster, the system never leaves that set because the played actions can

only *confirm* the belief in the self-confirming (n_1, n_2)-solution forecasting rule. More precisely, the system then cycles between the states in the self-confirming (n_1, n_2)-solution cluster according to the (cyclical) sequence of last $\text{Max}(n_1, n_2) + 2$ actions generated by the path of the associated self-confirming (n_1, n_2)-solution. The following Proposition establishes that the self-confirming (n_1, n_2)-solution clusters are the only absorbing sets of the process without trembles.

PROPOSITION 1. *The only absorbing sets of the learning process without trembles are the sets of states in* $\mathbb{E} \cup \mathbb{S}$.

5.2. *Stochastically Stable Sets*

We now proceed to analyze how the dynamics of the learning process is affected by the presence of rare trembles. For every $\varepsilon > 0$, we first note that the learning process defines an aperiodic dynamics because it is always possible to move from one state to another with an appropriate number of mistakes (and appropriate realizations of the random device in the updating process). From the theory of Markov processes, that property ensures that (1) the learning process has a unique stationary distribution, (2) the proportions of states reached along any sample path approach this distribution almost surely, and (3) the distribution of states at time t approaches this distribution as t gets large.

We thus obtain a unique stationary distribution for each probability of tremble ε. We study the limit of these stationary distributions as the probability of mistake ε gets small (keeping all other parameters fixed). The limit distribution is termed the *stochastically stable distribution.*

PROPOSITION 2. *The stochastically stable distribution places positive weight only on* (n_1, n_2)*-solution clusters* (*i.e., in* $\mathbb{E}$) *and no weight on self-confirming* (n_1, n_2)*-solution clusters that are not* (n_1, n_2)*-solution clusters* (*i.e., in* $\mathbb{S}$).

The technique involved in establishing this result relies on Freidlin and Wentzell (1984). Intuitively, with vanishing trembles the system spends virtually all of its time in absorbing sets of the learning process without trembles or equivalently the stochastically stable distribution allocates all of its probability to such sets. Transitions from one absorbing set to another can be accomplished only by means of trembles. The system will asymptotically allocate all of its probability to absorbing sets that are easy to reach and from which it is hard to get out. The proof of Proposition 2 consists in showing that it is much harder to go from the set $\mathbb{E}$ of (n_1, n_2)-solution clusters into the set $\mathbb{S}$ of-self-confirming (n_1, n_2)-solution clusters (that are not in $\mathbb{E}$) than the other way around.

The interpretation of Proposition 2 is thus that eventually players learn to play some (n_1, n_2)-solution as opposed to a self-confirming (n_1, n_2)-solution. It should be noted that Proposition 2 only guarantees that some

(n_1, n_2)-solution will emerge, but does not specify further which (n_1, n_2)-solution is more likely to emerge. In fact, it may well be that the stochastically stable distribution assigns a strictly positive probability only to a subset of the (n_1, n_2)-solution clusters as opposed to every (n_1, n_2)-solution cluster. Since the main purpose of this paper is to show that some (n_1, n_2)-solution will emerge, we do not go into that analysis, but in principle the (n_1, n_2)-solution concept could be refined on that basis. A *refined* (n_1, n_2)-solution would be such that it is reached with a strictly positive probability in the stochastically stable distribution.

The proof of Proposition 2 requires several steps, and introducing some new notation. First, for any state g in $\mathbb{G}$, we let $G(g)$ denote the set of states g' in $\mathbb{G}$ such that the states of belief associated with g and g' may differ *only* in their assignment of the players' plausibility levels of those forecasting rules which are not the most plausible one. That is, for any g' in $G(g)$, player i's most plausible forecasting rule according to g' coincides with player i's most plausible forecasting rule according to g, and the streams of $\text{Max}(n_1, n_2) + 2$ actions in g' and g are the same. Assume g is a (self-confirming) (n_1, n_2)-solution state. Then in the dynamics without trembles starting from states of belief in g or in any $g' \in G(g)$ yields the same sequence of plays (i.e., that induced by the associated (self-confirming) (n_1, n_2)-solution), since g and g' have the same most plausible forecasting rules which are always confirmed by the observations throughout the play. Second, we define the notions of sequences of trembles and of isolated trembles.

DEFINITION 8. (1) A *sequence* of trembles is a sequence $\{\delta^t\}_{t=1}^{\infty}$, where $\delta^t = 1$ if there is a tremble at period t and $\delta^t = 0$ if there is no tremble at t. It is referred to as an *infinite* sequence of trembles whenever the number of periods where $\delta^t = 1$ is infinite.

(2) A sequence of *isolated* trembles is a sequence of trembles $\{\delta^t\}_{t=1}^{\infty}$ such that there are at least $2\,\text{Max}(n_1, n_2)$ periods between two consecutive trembles. That is, if $\delta^t = 1$ then $\delta^{t+1} = 0$, $\delta^{t+2} = 0, \ldots, \delta^{t+2\,\text{Max}(n_1, n_2)} = 0$.

The following lemma is a key step in the proof of Proposition 2.

LEMMA 1. (i) *Starting from a (n_1, n_2)-solution state g, the system never leaves the set of states $G(g)$ whenever trembles occur according to a sequence of isolated trembles.*

(ii) *Starting from a self-confirming (n_1, n_2)-solution state that is not a (n_1, n_2)-solution state, there always exists a finite sequence of isolated trembles that leads with a strictly positive probability to a (n_1, n_2)-solution state.*

The intuition for Lemma 1 is as follows. Consider a self-confirming (n_1, n_2)-solution state that is not a (n_1, n_2)-solution state. As long as there are no trembles the system cycles between the states of the corresponding

self-confirming (n_1, n_2)-solution cluster. By assumption there must exist at least one time of the cycle induced by the self-confirming (n_1, n_2)-solution where the most plausible forecasting rule of a player, say player 1, yields an incorrect prediction off the equilibrium path. Assume that a tremble occurs at such a time, say at period t, and yields an action with an incorrect n_1-length prediction according to player 1's most plausible forecasting rule f_1. Assume also that no tremble occurs before $2\,\mathrm{Max}(n_1, n_2)$ periods. From period t to period $t + n_1 - 1$, the only effect on player 1 of the period t tremble is to increase the plausibility levels of some forecasting rules which previously had incorrect predictions on the equilibrium path (and had the prediction fitting with the tremble). It should be noted that during those periods player 1's most plausible forecasting rule remains the same (since the plausibility of forecasting rules having incorrect predictions on the equilibrium path prior to t can only reach the plausibility level $k_1 - n_1$ at best and $k_1 - n_1 > 1$). At period $t + n_1$, the period t predictions can be compared with the stream of realized actions from period $t + 1$ to period $t + n_1$. Since there were no trembles during those periods, the current most plausible forecasting rule f_1 of player 1 yields an incorrect prediction, and therefore becomes a little less plausible relative to any forecasting rule $f_1' \in F_1$ having correct predictions both on and off the equilibrium path (for this particular tremble). If such a forecasting rule f_1' happens to be as plausible as f_1 at period t, then the state of belief of player 1 is destabilized and may lead to any most plausible forecasting rule. Otherwise, after n_1 other periods without trembles, the system has returned to the original state of belief of player 1 except that some forecasting rules including f_1' have now a plausibility level that has reduced by one increment. Clearly after a finite number of such isolated trembles the self-confirming (n_1, n_2)-solution state will be destabilized, and may lead to some (n_1, n_2)-solution state.[11] It should be noted that the same argument would not apply if the most plausible forecasting rule f_1 originated from a (n_1, n_2)-solution rather than from a self-confirming (n_1, n_2)-solution. The reason is that now at period $t + n_1$ no forecasting rule other than f_1 would see its plausibility level decrease, as the period t prediction of f_1 would be correct.

The rest of the proof for Proposition 2 goes as follows. The first part of Lemma 1 shows that nonisolated (or consecutive) trembles are needed to destabilize a (n_1, n_2)-solution cluster. Proposition 2 is obtained by observing that consecutive trembles are far less likely than finite sequences of isolated trembles (occurring at given times of the cycle induced by a self-confirming (n_1, n_2)-solution).

[11] The complete argument requires looking at the state of belief of player 2 as well; see the Appendix.

5.3. *An Example*

The aim of this subsection is to provide a simple example illustrating the learning dynamics and its convergence properties. To this end, we construct an example with both a self-confirming (n_1, n_2)-solution and a (n_1, n_2)-solution, and we show how the system is likely to go from the former to the latter. Let $n_1 = n_2 = 1$, $A_1 = \{U, D\}$, $A_2 = \{L, R\}$. To simplify the exposition we will assume that F_1 consists of only two time-independent forecasting rules $f_1 = \{f_1^t\}$, $f_1' = \{f_1'^t\}$ such that $\forall t$, $f_1^t(U) = L$, $f_1^t(D) = R$, and $f_1'^t(U) = f_1'^t(D) = L$. The two rules differ in their predictions when D is played. Similarly, F_2 consists of two time-independent forecasting rules $f_2 = \{f_2^t\}$, $f_2' = \{f_2'^t\}$ such that $\forall t$, $f_2^t(L) = f_2^t(R) = U$ and $f_2'^t(L) = D$, $f_2'^t(R) = U$, which differ in their prediction when L is played. We assume that $\nu_1(LUL) > \nu_1(LDR)$ and $\nu_2(ULU) > \nu_2(URU)$ so that the forecasting rule profile (f_1, f_2) is that of a self-confirming (n_1, n_2)-solution (on the equilibrium path the actions U and L are played). We also assume that $\nu_1(LDL) > \nu_1(LUL)$, $\nu_1(RUL) > \nu_1(RDL)$ and $\nu_2(ULD) > \nu_2(URU)$, $\nu_2(DLD) > \nu_2(DRU)$ so that the profile (f_1', f_2') is that of a (n_1, n_2)-solution (the equilibrium path is $DLDL\ldots$). We assume that $\nu_2(DLU) > \nu_2(DRU)$ so that (f_1, f_2) does not correspond to a (n_1, n_2)-solution (if the off equilibrium path action D is played, player 2 prefers action L to action R given his forecast and this contradicts player 1's forecast). Finally, we assume that players use three levels of plausibility: $k_1 = k_2 = 3$.

Start from the global state that is most favorable to the self-confirming (n_1, n_2)-solution generated by (f_1, f_2). That is, the initial states of belief satisfy $s_i(f_i) = 1$, $s_i(f_i') = 3$, for $i = 1, 2$, and the last actions of the initial state correspond to the path induced by the self-confirming (n_1, n_2)-solution (f_1, f_2). As long as there are no mistakes the actions U and L are played by players 1 and 2, respectively, and their states of belief remain unchanged. Assume that at some period t player 1 makes a mistake and plays D. Then the new state of belief of player 2 is $s_2^t(f_2) = 1$, $s_2^t(f_2') = 2$ (f_2' becomes a little more plausible as $f_2'^{t-1}(L) = D$ and $f_2^{t-1}(L) = U$). Given that player 2's most plausible forecasting rule is f_2, player 2 selects the action L (if he does not make a mistake) at period $t + 1$ because $\nu_2(DLf_2^t(L)) > \nu_2(DRf_2^t(R))$. That action of player 2 makes the forecasting rule f_1 a little less plausible (since $f_1^t(D) = R$) and the forecasting rule f_1' a little more plausible (since $f_1'^t(D) = L$) to player 1. That is, the new state of belief of player 1 is $s_1^{t+1}(f_1) = 1$, $s_1^{t+1}(f_1') = 2$. At period $t + 2$, player 1 chooses U (if he does not make a mistake) since f_1 is still the most plausible forecasting rule for him and L has just been played. It follows that the period $t + 2$ state of belief of player 2 is again s_2, i.e., $s_2^{t+2}(f_2) = 1$, $s_2^{t+2}(f_2') = 3$. In later periods $t' > t$, as long as there are no

mistakes the sequence of plays is $ULUL\ldots$, and the states of belief of players 1 and 2 are, respectively, $s_1^{t'-1}(f_1) = 1$, $s_1^{t'-1}(f_1') = 2$ and $s_2^{t'}(f_2) = 1$, $s_2^{t'}(f_2') = 3$. Assume that at some period t' player 1 makes again a mistake and plays D. Player 2's period t' state of belief is $s_2^{t'}(f_2) = 1$, $s_2^{t'}(f_2') = 2$, and player 2 selects L at period $t' + 1$. This implies that $\bar{s}_1^{t'+1}(f_1) = \bar{s}_1^{t'+1}(f_1') = 2$, and therefore player 1's state of belief may switch to any state of belief in particular to $s_1^{t'+1}$, where $s_1^{t'+1}(f_1) = 3$, $s_1^{t'+1}(f_1') = 1$. In such a case, player 1 chooses U at period $t' + 2$ (when he does not make a mistake), since $\nu_1(LDf_1'^t(D)) > \nu_1(LUf_1'^t(U))$ and player 2 has played L in period $t' + 1$. Then $\bar{s}_2^{t'+2}(f_2) = \bar{s}_2^{t'+2}(f_2') = 2$, and player 2's state of belief is destabilized so that it may switch to $s_2^{t'+2}$, where $s_2^{t'-2}(f_2) = 3$, $s_2^{t'-2}(f_2') = 1$. From then on, f_i' remains player i's most plausible forecasting rule for player $i = 1, 2$ as long as there are not consecutive trembles, and therefore the (n_1, n_2)-solution path $(DLDL\ldots)$ generated by (f_1', f_2') is played.

6. DISCUSSION

The learning process involves several elements of bounded rationality (in addition to the feature of limited forecasting). We wish now to discuss the role and interpretation of each assumption.

6.1. *On the Finiteness of the Number of Plausibility Levels*

The state of belief as defined in Section 4 reflects some limited capability of the players in their information treatment, where the limitation bears on two points. First the updating of the state of belief relies only on the information whether the predictions of forecasting rules $f_i \in F_i$ coincide with the observation or not as opposed to how many mistakes are required to explain the observation with f_i. Second, each player i restricts himself to a finite number k_i of plausibility levels as opposed to a potentially larger (or infinite) number. The first limitation is not crucial, and there would be no conceptual difficulty in making the updating depend on the number of mistakes required to explain the observation with each forecasting rule $f_i \in F_i$. (We have made it for notational purposes.) The second limitation is more essential. Correct Bayesian updating would require a perfect record of how many mistakes are needed to explain past observations with each forecasting rule in the support of belief. Thus, for t large enough, with the standard Bayesian view, the plausibility level of some forecasting rule could go beyond k_i. It should be noted though that if one accepts that player i restricts himself to k_i plausibility levels, the kind of updating proposed in Section 4 seems reasonable (in that it is as close as possible to the correct Bayesian updating).

Technically, the *bound* on the number of plausibility levels allows us to reduce the learning process to a Markov process with a *finite* state space. This is used to apply the techniques of perturbed Markov processes in Section 5. On an interpretative level, the bound on the number of plausibility levels seems a reasonable way to capture the idea that a boundedly rational player is likely to form his belief on the basis of some *imperfect* record of past observations.[12] We wish to stress that our finding that eventually players learn to play a (n_1, n_2)-solution is robust to any change in the bound on the number of plausibility levels as long as $k_i > n_i + 1$.[13] In this sense, the limit behavior is not too sensitive to the degree of bounded capability in information treatment.

6.2. *On the Uniqueness of the Most Plausible Forecasting Rule*

We have assumed that at each point of the process the players have a uniquely defined most plausible forecasting rule (which, of course, may change from period to period). We wish to point out that the above analysis can easily be extended to the case where the players may have several most plausible forecasting rules at the same time still assuming though that when a player has several most plausible forecasting rules there is a chance (which may be arbitrarily small) that his state of belief switches to any conceivable state of belief (not necessarily one with a uniquely defined most plausible forecasting rule). Specifically, the learning framework would then be adapted as follows: Each forecasting rule $f_i \in F_i$; would have an a *a priori* weight denoted $\mu_i(f_i)$. Whenever there is no tremble at period t the current action a_i^t would be chosen so as to maximize $\sum_{f_i \in S_i^t(1)} \mu_i(f_i)\nu_i(a_j^{t-1} a_i f_i^t(a_i))$, where a_j^{t-1} is the period $t-1$ action and the sum bears over all period t most plausible forecasting rules $f_i \in S_i^t(1)$ (i.e., such that $s_i^t(f_i) = 1$). Finally, the updating of players' states of belief (including the possibility of switch) would be defined exactly in the same fashion as in Section 4. With easy adaptations of the proofs, it can be shown that the learning framework just described yields the same asymptotic results (Propositions 1 and 2) as the one described above. The framework with uniquely defined most plausible forecasting rules was chosen mostly for notational purposes.

[12] The limitation imposed by k_i in the state-of belief does not reduce to bounded recall (where the players remember only a finite number of past actions). If imperfect record were to take the form of bounded recall, then the result of Proposition 2 would only hold if the memory capacity of the players increases to infinity at the same time as the probability of making a mistake goes to zero (the complete argument would be significantly harder to establish).

[13] We have used that assumption to guarantee that forecasting rules which have incorrect predictions on the equilibrium path have a sufficiently high plausibility level.

6.3. *On the Random Device When There Are Several Most Plausible Forecasting Rules*

We have assumed that whenever a player has several forecasting rules which may be the most plausible one, there is a chance (which may be arbitrarily small) that the player switches to any conceivable state of belief, which means that any forecasting rule may become the most plausible one. On a technical level, we have used that assumption to ensure that the only absorbing sets of the process without mistakes are the self-confirming (n_1, n_2)-solution clusters. (Otherwise, it would *a priori* be conceivable that the process without mistakes admit other absorbing sets in which the most plausible forecasting rule of a player varies in a cyclical fashion, say.) Whether that assumption is needed for Proposition 2 is an open (and presumably difficult) question left for future research. (Note that in the example, we would obtain the same conclusion if the assumption were to be dropped because there are only two forecasting rules.)

The random device assumption can again be interpreted in terms of bounded rationality, and it should be noted that the convergence result to (n_1, n_2)-solutions does not depend on the exact specification of the distribution of change in the state of belief. It is nevertheless probably the least satisfactory feature in the learning process, and it would be of interest to analyze what happens asymptotically to the process if that assumption were to be dropped. A possible defense for the assumption is as follows. Given that player i's learning bears on forecasting rules, player i may be thought of as being mostly concerned with the determination of the most plausible forecasting rule. When there is exactly one most plausible forecasting rule, player i is satisfied. When there are two or more forecasting rules that can be the most plausible one, player i is unhappy (or feels that something is wrong with his belief) and must change something. It seems then reasonable to model that change as a switch to any conceivable state of belief (where the switch is assumed not to be fully under the control of the player).[14, 15]

[14] There are other ways to model the disenchantment of player i. For example, when the previous most plausible forecasting rule becomes as likely as some other forecasting rule, then the previous most plausible forecasting rule may be assigned to the set of least plausible forecasting rules with positive probability. Such a specification would yield the same asymptotic results (with a slightly more complicated proof).

[15] Another interpretation of the state of belief perturbation is that the player misperceives with positive probability the plausibility levels of the forecasting rules other than the most plausible one, and therefore when there are several most plausible forecasting rules the state of belief may switch to any conceivable state. In establishing Proposition 2, that interpretation would, however, require that the probability of misperception goes to zero together with the probability of trembling, which would complicate the argument.

6.4. *On the Mistakes*

In the learning process we have assumed that each player i could make a mistake at each period where he must move. Also, the distribution of mistakes was assumed to be independent from period to period and stationary. Given the analysis of Section 5 it should be clear that the feature of the mistake distribution that drives the convergence result is that consecutive trembles are far less likely than isolated ones (in the sense made precise in Definition 8). This is because isolated trembles are sufficient to destabilize a self-confirming (n_1, n_2)-solution (that is not a (n_1,n_2)-solution) cluster, whereas consecutive trembles are required to destabilize a (n_1, n_2)-solution cluster. The asymptotic result of Section 5 would thus continue to hold as long as that property is met whether or not the distributions of mistakes are assumed to be stationary and independent from period to period (on state-dependent perturbations, see also Bergin and Lipman, 1996).

6.5. *On the Forecasting Rule*

Throughout the paper, each player i was assumed to make his choice of action on the basis of his forecast over the forthcoming n_i moves (including his own moves). In some cases, the forecast can be thought of as bearing only on the reaction function of the other player over the forthcoming n_i moves. In such cases, one might argue that player i makes a *plan* of actions over the forthcoming n_i moves given his forecast about the reaction function of the other player. His plan of actions leads him in turn to choose a current action, which is the effective choice to be made in the current period. At every period where player i must move, player i would then make plans of actions over the forthcoming n_i periods yielding a choice of current action. For such a process of thought, it may well be that his effective choices of action in the next periods do not coincide with the plan originally made even though player i has a correct forecast about the reaction function of the other player. In other words, this process of thought might lead to time inconsistencies. In the long run, if player i learns the reaction function of the other player he should realize that his plans of actions do not coincide with the actions he effectively chose in the next periods. The only way for player i to avoid time inconsistencies is to reduce his current period choice to his current period action. The formulation adopted in this paper is the only one compatible with that view.

7. CONCLUSION

A learning process was proposed in which players eventually learn to play a (n_1, n_2)-solution. The process involved several elements of bounded

rationality including some which are not linked to the limited forecasting form of bounded rationality. Still, as suggested in the discussion (Section 6) the convergence to (n_1, n_2)-solution is robust to a number of variations of how the other sorts of bounded rationality are modelled as long as the players are assumed to keep the same length of foresight throughout the process. A possible extension would allow player i to change his length of foresight when he feels that he understands sufficiently well the forthcoming n_i moves as a function of his own move. He might then (potentially) decide to increase his length of foresight from n_i to $n_i + 1$, say. In this view, starting from the (n_1, n_2)-solution that is currently played, player i might infer from the observed sequence of forthcoming $n_i + 1$ actions new $(n_i + 1)$-length forecasts. In the case this induces a modification of his behavior, a second stage of learning would lead to some $(n_i + 1, n_j)$-solution. Otherwise, the original (n_1, n_2)-solution was already a $(n_i + 1, n_j)$-solution. Such a process of changes of lengths of foresight might be pursued for both players defining a new stochastic process. If there exist strategy profiles that define a (n_1, n_2)-solution for all (n_1, n_2) sufficiently large (such strategy profiles are termed hyperstable solutions in Jéhiel, 1995), then these are absorbing states of the overall process, and one might conjecture that they will emerge in the long run. The precise analysis is left for future research.[16]

APPENDIX

Proof of Proposition 1. It is rather immediate to check that sets in $\mathbb{E} \cup \mathbb{S}$ are absorbing. To see this consider a self-confirming (n_1, n_2)-solution state, and observe that when there are no trembles the played actions always confirm the prediction of the self-confirming (n_1, n_2)-solution forecasting rule. As a result the self-confirming (n_1, n_2)-solution forecasting rule remains the most plausible forecasting rule all along the played path Moreover, those forecasting rules which give the correct predictions along the played path keep the same plausibility level and those which give some incorrect predictions have a plausibility level set to k_i.

Conversely, consider an absorbing set. We wish to show that it is necessarily a self-confirming (n_1, n_2)-solution cluster. If the same forecasting rule remains the most plausible one for each player i all along the learning process generated by a state in the absorbing set, then it is rather straightforward to see that the absorbing set must be a self-confirming

[16] It might as well be argued that, when a player feels that he understands sufficiently well the sequence of forthcoming actions, he decides to be less sophisticated and reduces his length of foresight by one increment, say. As long as the length of foresight remains above the threshold defined by the hyperstable solution the play is unaffected.

(n_1, n_2)-solution cluster. If the most plausible forecasting rule does change for at least one player, say player 1, along the learning process generated by some state in the absorbing set, then it must be that the random device in the updating process of player 1's state of belief is triggered at some points in time. Consider such a period t. There is a positive probability that the next state of belief of player 1 corresponds to that of a self-confirming (n_1, n_2)-solution state. If player 2's state of belief corresponds to that same self-confirming (n_1, n_2)-solution state, then the Markov process has reached a self-confirming (n_1, n_2)-solution cluster which is absorbing. It will thus never return to the original absorbing set, yielding a contradiction. If player 2's state of belief does not correspond to that self-confirming (n_1, n_2)-solution state, then eventually (after a finite number of periods) some forecasting rule of player 2 other than the most plausible at period t will become equally plausible. At such a time the random device of the updating of player 2 will be triggered and there is a positive probability that player 2's new state of belief now corresponds to the self-confirming (n_1, n_2)-solution state.[17] From then on, the system will never leave the associated self-confirming (n_1, n_2)-solution cluster, and therefore will never return to the original absorbing set, a contradiction. Q.E.D.

Proof of Lemma 1. (1) Consider a (n_1, n_2)-solution state. Assume player 1 makes a mistake at period t, and that there is no further tremble up to period $t + 2\,\mathrm{Max}(n_1, n_2)$. We first analyze the effect of such a tremble on player 1's state of belief. Compared to the description in the main text, the only difference is that, at period $t + n_1$, the current most plausible forecasting rule of player 1 yields a correct prediction, and therefore the updating can only reinforce player 1's belief in his most plausible forecasting rule (the plausibility level of those forecasting rules which yielded incorrect predictions after the period t tremble will increase by one increment if possible). Regarding the effect (of player 1's tremble) on player 2, we note that it is only temporary and after $2n_2$ periods, player 2's state of belief returns to his original period t state of belief. Clearly, the argument shows that infinite sequences of isolated trembles are unable to destabilize a (n_1, n_2)-solution cluster.

(2) Consider a self-confirming (n_1, n_2)-solution state that is not a (n_1, n_2)-solution state. As explained in the main text, consider a time of the cycle induced by the self-confirming (n_1, n_2)-solution where the most plausible forecasting rule of a player, say player 1, yields an incorrect prediction off the equilibrium path. Assume that a tremble occurs at such a time, say at period t, and yields an action with an incorrect n_1-length

[17] If in the meantime the random device of player 1 is triggered there is always a positive probability that he returns always to the same self-confirming (n_1, n_2)-solution state of belief.

prediction (with no tremble before $2\,\mathrm{Max}(n_1, n_2)$ periods). The effect on player 1's subsequent states of belief has been described in the main text. If the updating at period $t + n_1$ yields several most plausible forecasting rules to player 1 then his state of belief will switch. If the updating at period $t + n_1$ yields a uniquely defined most plausible forecasting rule, then there is no random switch in player 1's state of belief at period $t + n_1$, but after period $t + 2n_1$ the plausibility level of some forecasting rules (yielding the correct predictions on the equilibrium path and after the period t tremble) has reduced by one increment. Moreover as noted earlier the effect (of player 1's, tremble) on player 2 is only temporary and after $2n_2$ periods, player 2's state of belief has returned to his original period t state of belief (there is no experimentation for player 2). Clearly, after at most $k_1 - 1$ such isolated trembles, there will be a period where (in the updating process) player 1 has more than one plausible forecasting rule. Player 1's state of belief may then switch to any state of belief, in particular to a state s_1 such that $s_1(f_1') = 1$, $s_1(f_1) = k_1$ for every forecasting rule $f_1 \neq f_1'$, where (f_1', f_2') is the forecasting rule profile of a (n_1, n_2)-solution for some (player 2's forecasting rule) f_2'. Let f_2 be player 2's most plausible forecasting rule at the time of the switch. Lemma 1 follows if (f_1', f_2) is the forecasting rule profile of some (n_1, n_2)-solution. If (f_1', f_2) is not the forecasting rule profile of a (n_1, n_2)-solution, then two cases may arise. Either (f_1', f_2) is the forecasting rule profile of a subjective (n_1, n_2)-solution or not. When (f_1', f_2) does not correspond to a subjective (n_1, n_2)-solution, then player 2's state of belief will be destabilized even without further trembles. That is, after a finite number of periods (without trembles), player 2's updating will result in several most plausible forecasting rules. When (f_1', f_2) does correspond to a subjective (n_1, n_2)-solution which is not a (n_1, n_2)-solution, then by the same argument displayed for player 1, one can show that player 2's state of belief will be destabilized after a finite number (at most $k_2 - 1$) of player 2's isolated trembles.[18] When player 2's state of belief is destabilized, there is a chance that his state of belief switches to s_2, where $s_2(f_2') = 1$, and $s_2(f_2) = k_2$ for other forecasting rules, where f_2' is player 2's forecasting rule that has been introduced above (i.e., (f_1', f_2') defines a (n_1, n_2)-solution). We have thus shown that a finite number (at most $k_1 + k_2 - 2$) of isolated trembles may lead with positive probability to a (n_1, n_2)-solution state. Q.E.D.

Proof of Proposition 2. Proposition 2 follows from Proposition 1, Lemma 1, the general analysis of perturbed Markov processes (see Freidlin and Wentzell, 1984), and the observation that when the probability of tremble

[18] If player 1's state of belief is destabilized in the meantime, his state of belief may go back to s_1 as defined above.

goes to zero a finite sequence of isolated trembles (occurring at given times of a cycle of finite length) is infinitely more likely than consecutive trembles over a number $T \approx N/\varepsilon$ of periods (where $N = k_1 + k_2 - 2$ is the maximum number of isolated trembles required to destabilize a self-confirming (n_1, n_2)-solution state, see above). The reason is that, over $T = N/\varepsilon$ periods, the expected number of trembles is N, and as ε goes to zero the probability that there be consecutive trembles becomes negligible as opposed to the probability of having N isolated trembles destabilizing some given self-confirming (n_1, n_2)-solution state. Q.E.D.

REFERENCES

Abreu, D., and Rubinstein, A. (1988). ''The Structure of Nash Equilibrium in Repeated Games with Finite Automata,'' *Econometrica* **56**, 383–396.

Bergin, J., and Lipman, B. L. (1996). ''Evolution with State-Dependent Mutations,'' *Econometrica* **64**, 943–956.

Binmore, K., and Samuelson, L. (1992). ''Evolutionary Stability in Repeated Games Played by Finite Automata,'' *J. Econ. Theory* **57**, 278–305.

Foster, D., and Young, H. P. (1990). ''Stochastic Evolutionary Game Dynamics,'' *Theoret. Popul. Biol.* **38**, 219–232.

Freidlin, M., and Wentzell, A. (1984). *Random Perturbations of Dynamical Systems*. New York: Springer-Verlag.

Fudenberg, D., and Harris, C. (1992). ''Evolutionary Dynamics in Games with Aggregate Shocks,'' *J. Econ. Theory* **57**, 420–441.

Fudenberg, D., and Levine, D. (1993). ''Self Confirming Equilibrium,'' *Econometrica* **61**, 523–545.

Jéhiel, P. (1995). ''Limited Horizon Forecast in Repeated Alternate Games,'' *J. Econ. Theory* **67**, 497–519.

Jordan, J. S. (1991). ''Bayesian Learning in Normal Form Games,'' *Games Econ. Behav.* **3**, 60–81.

Kalai, E., and Lehrer, E. (1993a). ''Rational Learning Leads to Nash Equilibrium,'' *Econometrica* **61**, 1019–1046.

Kalai, E., and Lehrer, E. (1993b). ''Subjective Equilibrium in Repeated Games,'' *Econometrica* **61**, 1231–1240.

Kalai, E., and Stanford, W. (1988). ''Finite Rationality and Interpersonal Complexity in Repeated Games,'' *Econometrica* **56**, 397–410.

Kandori, M. (1996). ''Evolutionary Game Theory in Economics,'' paper presented at an invited symposium of the Seventh World Congress of the Econometric Society (August 1995, Tokyo).

Kandori, M., Mailath, G., and Rob, R. (1993). ''Learning, Mutation and Long Run Equilibria in Games,'' *Econometrica* **61**, 29–56.

Kandori, M., and Rob, R. (1995). ''Evolution of Equilibria in the Long Run: A General Theory and Applications,'' *J. Econ. Theory* **65**, 383–414.

Lehrer, E. (1988). ''Repeated Games with Stationary Bounded Recall Strategies,'' *J. Econ. Theory* **46**, 130–144.

Nachbar, J. H. (1997). ''Predictions, Optimization and Learning in Repeated Games,'' *Econometrica* **65**, 275–309.

Neyman, A. (1985). ''Bounded Complexity Justifies Cooperation in Finitely Repeated Prisoner's Dilemma,'' *Econ. Lett.* **19**, 227–229.

Nöldeke, G., and Samuelson, L. (1993). ''An Evolutionary Analysis of Backward and Forward Induction,'' *Games Econ. Behav.* **5**, 425–454.

Rubinstein, A. (1986). ''Finite Automata Play the Repeated Prisoner's Dilemma,'' *Econ. Theory* **39**, 83–96.

Selten, R. (1975). ''Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games,'' *Int. J. Game Theory* **4**, 22–55.

Young, H. P. (1993a). ''The Evolution of Conventions,'' *Econometrica* **61**, 57–84.

Young, H. P. (1993b). ''An Evolutionary Model of Bargaining,'' *J. Econ. Theory* **59**, 145–168.