False Reputation in a Society of Players

Matthew O. Jackson and Ehud Kalai *

March 1997

Abstract

Exploiting small uncertainties on the part of opponents, players in long, finitely repeated games can maintain false reputations that lead to a large variety of equilibrium outcomes. Even cooperation in a finitely repeated prisoners' dilemma is obtainable. Can such false reputations be maintained in a society if the same repeated game is played recurringly by many different groups and each group observes the play paths of the earlier groups? We argue that such false reputations must die out over time. To prove this in environments that allow for rich (uncountable) sets of types of players, we combine ideas of purification with recent results from the rational learning literature.

*M.E.D.S., Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208-2009. We thank Roger Myerson for a conversation out of which this project was born. We gratefully acknowledge financial support under grants SBR 9515421 and SBR 9507912 from the National Science Foundation.

1 Introduction

The role of reputation in strategic interaction is a topic of major concern to economists, and has received considerable attention in the literature. Kreps, Milgrom, Roberts, and Wilson (1982) (KMRW for short) show that even the famous paradox of the finitely repeated prisoners' dilemma game disappears when reputation phenomena are brought into the analysis. A small uncertainty about a player's preferences can be used by the player to create a favorable false reputation in a long game. Such a false reputation can lead to long periods of cooperation in equilibrium.

Although the details are complicated, the intuition behind the KMRW result is straightforward: given a small chance that a player's opponent may have a natural propensity to cooperate (in particular, play tit-for-tat so that cooperation begets cooperation and finking begets finking) and, given a long enough horizon, a player is willing to risk cooperation anticipating that it may lead to longer periods of cooperation. The opponent who may not in fact have this natural propensity for cooperation is then faced with a choice: he can fink and take advantage of the other player's cooperative play this period, but thereby destroy his reputation and face finking thereafter, or he can cooperate, thereby not allowing the other player to distinguish his true type. By cooperating, the opponent maintains the other player's uncertainty and his "false reputation," provided the remaining horizon is long enough to warrant chance-taking by the other player. Eventually the horizon is so short that even in the face of the maintained uncertainty the player will no longer be willing to chance cooperating; and so the opponent no longer has an incentive to maintain his "false reputation" and both players fink.

Generalizing this result, Fudenberg and Maskin (1986) establish a reputation "folk" theorem. They study an n-person normal form game G played repeatedly m times with perfect monitoring ($m \times G$, for short) and show that if there is a small uncertainty about players' payoffs, but the number of repetitions m is large, then any vector of individually rational payoffs of the stage game G is obtainable as the average payoff of a Bayesian equilibrium of a Bayesian version of $m \times G$.

While the above papers show that false reputations can be sustained for significant periods of time in a sufficiently long finitely repeated games, it is not clear that the same phenomenon can be sustained by an entire society. If many different groups of players play $m \times G$ recurringly, and later groups of players observe the play of earlier groups, does the ability to maintain false reputation disappear?

We study this question in a Bayesian recurring repeated game model where the stage game is an $m \times G$ game. (To avoid confusion with a double use of the term stage game, the m iterations within a stage game are called rounds.) The recurring version of this game is played in periods $t = 1, 2, ...$ as follows. In period 1 a group of n-players play $m \times G$ and the resulting play path becomes public knowledge available to all future players. In period 2 a new group of n-players play $m \times G$ and their resulting play path becomes publicly known. This iteration continues indefinitely so that after every social history of play paths a new group of n-players play the stage game and add one play path to the cumulative social history.

Each of the infinitely many players in the recurring game has preferences over the play paths of his or her own stage game. To allow for uncertainty about opponents' preferences, an essential ingredient for the reputation question, the recurring game is augmented to be a Bayesian Recurring game. In a preliminary stage, according to a commonly known prior probability distribution, an infinite vector of Harsanyi (1967) types is drawn and each player in the infinite horizon is informed of his or her own realized preferences. With this private information they proceed to play the infinite recurring game described above.

Since the first stage of the Bayesian recurring game is the Fudenberg and Maskin game, initial uncertainty can lead to a large variety of reputations and possible play paths. For example, if the underlying normal form game G is a prisoners' dilemma game and all the players in the Bayesian recurring game were selected to have standard prisoners' dilemma preferences, but the prior distribution assigns positive probability to player types that prefer to use a tit-for-tat strategy, then (following the results of KMRW) the first play path will be mostly cooperative. It is not clear, however, what late play paths are likely to be. There seem to be two competing arguments.

One argument, combining learning and backwards induction, builds on the fact that all of the realized types hold the prisoners' dilemma preferences and so they will certainly fink in the last round of their stage games. After learning to predict this, later players will fink in second to last round. As later players learn to predict this, they will fink in the third to last round, and so on.

The other argument leads to an opposite view. After learning to fink in the first round, a cooperative first round action by a late player is likely to be interpreted by his opponent as a signal that the player is not a standard prisoners' dilemma type. In turn, this could help this player build a favorable reputation for the remaining $m - 1$ rounds of his stage game. Can we construct, for example, a cyclical Bayesian equilibrium of the recurring game where the play in different stages alternates between long periods of cooperation and long periods of finking? The answer to this question would seem to be yes, especially if the set of possible types is rich enough to allow for preferences of later players potentially to be unrelated to those of earlier players.

This paper illustrates the strength of the learning argument by showing that even when the prior probability distribution over types is (uncountably) rich, the false reputation phenomenon dies out with time and in equilibrium late players play near the equilibrium of the stage game. For example, in the Bayesian recurring repeated prisoners' dilemma game, for a large set of initial beliefs, if the realized players are close to the standard prisoners' dilemma types, then the probability of a finking play path approaches one over time. In more general terms, while for the first period of the Bayesian recurring repeated game one obtains a general folk theorem, in later periods one obtains what may be thought of as an *anti-folk* theorem.

To demonstrate this conclusion, this paper combines the purification ideas of Harsanyi (1973) with results from the recent literature on rational learning (relating to the merging of measures as shown by Blackwell and Dubins (1962) and Kalai and Lehrer (1993a)). Recent criticisms of the Kalai and Lehrer (1993a) approach center on the assumption that the prior assigns positive probability [1] to single vectors of types. Nachbar (1996) [2], building on an intuition of Jordan (1993), argues that if the set of types is sufficiently rich, then it is impossible to have players' predictions eventually become accurate, and to have the players best respond to those beliefs. The idea hinges on the fact that in a rich enough type space best responses generically are pure strategies, yet equilibrium convergence may require mixed strategies. The requirement of predictability cannot be met as it implies prediction of the actual pure strategy played, contradicting the best response requirement for a game with a mixed strategy equilibrium.

[1] Similar criticisms apply to the weaker assumption of absolute continuity.

[2] See also Foster and Young (1996).

We overcome this difficulty in the face of a rich type space by studying the play of ε-neighborhoods of a given player. We refer to the players in the neighborhood as ε-variants of the given player. Even though the prior probability of a given type may be zero, the prior probability on the player being some ε-variant may be positive. Our results rest on the fact that, although any single player will always be choosing pure strategies, *any* set of ε-variants of the player eventually will choose actions that average to the correct mixed strategy. Thus, Harsanyi's (1973) purification idea arises naturally and plays an important role in our recurring context.

The indirect contributions of this paper to rational learning are similar to, yet different from, some existing papers. First, like Jordan (1991), Nyarko (1994), and Lehrer and Smorodinsky (1997), this paper is restricted to learning Nash equilibrium within the play of Bayesian equilibrium. However, those papers study the repeated play of a one shot game and assume that the stage game actions are observed. [3] Given that our stage games are themselves repeated games, it is unreasonable to assume that the actions chosen in the stage game are observed, since those actions include prescriptions for play at potentially unreached nodes of the corresponding extensive form. Thus, we assume that only play paths are observed in the stage games, which means that the results mentioned above cannot be applied.

In addition to the weakening of the requirement of observability of stage game play, our assumptions concerning the prior probability differ from previous papers that have obtained results concerning convergence of equilibrium *play*. Lehrer and Smorodinsky (1997) and Sandroni (1997) assume a positive prior probability of all neighborhoods of the true type to draw conclusions about eventual equilibrium play. However, while the assumption in those papers relate to players' behaviors, the assumptions in our paper are on the primitive of preferences. [4]

The paper proceeds as follows: Section 2 presents the model, Section 3 presents the general theorem, Section 4 presents the proof of the theorem, Section 5 examines the repeated prisoners' dilemma and shows that a

[3] Jackson and Kalai (1995ab) allow for more general stage games. However, they assume countable sets of types and some relationship between payoffs and observability.

[4] Jordan (1991) and Nyarko (1994) also have measures over preferences as the primitive, but they obtain results concerning convergence of players' expectations, rather than play, and for the case where the stage game is a normal form game. The reader is referred to Marimon (1995) for a general survey of the learning literature.

strengthening of the theorem is possible, and Section 6 concludes.

2 The Model

The Stage Game

There are n *player* roles.

The stage game consists of m *rounds* ($m \geq 1$) of a finite normal form game played repeatedly by a fixed group of n players. We refer to iterations inside a stage as rounds, to distinguish them from stages.

A_i, with generic element a_i, is a finite set of possible *actions* that player i has available in the normal form game. Let $A = \times_i A_i$.

The possible results of the stage game are complete *play paths*, i.e., vectors $p = (a^1, a^2, ..., a^m)$ with $a^k \in A$ denoting the vector of actions taken at round k. A^m denotes the set of all possible play paths.

S_i, with generic element s_i, is a set of possible *strategies* that player i can use for playing the stage game. Following Kuhn's (1953) theorem, we consider mixed strategies of the extensive form stage game.

g_s denotes a probability distribution over play paths for a given vector of strategies $s = (s_1, \ldots, s_n)$.

The Recurring Game

The stage game is played sequentially but each time by a new group of n players. After each stage the play path resulting in the stage becomes publicly known (i.e., known to all future players).

For each time $t \in \{1, 2, 3, \ldots\}$ h^t denotes the *history* of play paths that resulted in stages *preceding* stage t. Thus, h^t is a vector $(p^1, \ldots, p^{t-1})$. We follow the convention that $h^1 = \emptyset$.

For each t the set of histories h^t is denoted H^t, and the set of all finite histories is $H = \cup_t H^t$.

After every history a new group of n players is selected to play the stage game. The pair (i, h^t) denotes the *player* who plays in role i at stage t if the history through t is h^t.

Types

The set denoting the possible *types* for players in role i is V_i, with generic element $v_i : A^m \rightarrow [0, 1]$.[5] A type v_i specifies the payoff to the player in role

[5]The bound on utility is not important to our results. It simply allows us to easily define a metric on the space and avoid using topological arguments.

i as a function of the play path played in his stage, and is assumed to satisfy the von Neumann-Morgenstern axioms. The type of player (i, h^t) is denoted u_{i,h^t}, where $u_{i,h^t} \in V_i$. $V \equiv \times_i V_i$ denotes the set of vectors of stage game types.

A *type profile for players in role* i (i-type profile, for short), is an infinite vector $u_i = (u_{i,h^t})_{h^t \in H}$ which specifies a type for player (i, h^t) for each $h^t \in H$.

A *social type profile* is a vector $u = (u_i)$ specifying an i-type profiles for every i. The set of i-type profiles is denoted U_i, and the set of social type profiles is denoted U.

The Bayesian Recurring Game

A *Bayesian recurring game* allows for uncertainty over the types of opponents as follows.

The Prior Distribution

Given two i-type profiles $u_i \in U_i$ and $u_i' \in U_i$, define the metric

$$|u_i - u_i'| = \sup_{t,h^t,p^t} |u_{i,h^t}(p^t) - u'_{i,h^t}(p^t)|.$$

Given social type profiles $u \in U$ and $u' \in U$, define the metric

$$|u - u'| = \max_i |u_i - u_i'|.$$

For $\varepsilon > 0$ let u^ε denote the neighborhood

$$u^\varepsilon = \{u' : |u - u'| \leq \varepsilon\}.$$

D_i is a *type-generating measure* over i-type profiles, i.e., a Borel probability measure over the set U_i. The support of D_i may be uncountable. D is the product measure, $D = \times_i D_i$, over the set of social type profiles U.

The social type profile u is *non-isolated* if $D(u^\phi) > 0$ *for all* $\phi > 0$.

Strategies

A *strategy* for player (i, h^t) is a Borel measurable function $\sigma_{i,h^t} : V_i \to S_i$, which specifies a stage game (mixed) strategy for (i, h^t) as a function of (i, h^t)'s type u_{i,h^t}.

A *social strategy profile* σ is an infinite vector of strategies $\sigma = (\sigma_{i,h^t})_{i,h^t}$, which specifies a strategy σ_{i,h^t} for each player (i, h^t).

The Bayesian recurring game is played as follows:

In an initial stage a social type profile u is chosen according to the prior D and each player (i, h^t) is informed of his or her own realized u_{i,h^t}.

Stage 1: Players $(1, h^1), \ldots, (n, h^1)$ select strategies as a function of their types, i.e., $s_i = \sigma_{i,h^1}(u_{i,h^1})$, to play the stage game. A play path p^1 is determined by the distribution g_s and the social history at the beginning of stage 2, $h^2 \equiv (p^1)$, is publicly revealed.

The infinite recurring game is defined inductively for $t = 2, 3, \ldots$.

Stage t: The new players $(i, h^t)_{1 \leq i \leq n}$ select stage-game strategies (s_i) as a function of their types, i.e., $s_i = \sigma_{i,h^t}(u_{i,h^t})$. A play path p^t is randomly selected by the distribution g_s and the social history of length t, $h^{t+1} \equiv (p^1, p^2, ..., p^t)$, becomes publicly known.

Outcomes and Probability

A fully specified *outcome* of the Bayesian recurring game is an infinite sequence

$$o = (u, p^1, p^2, ...).$$

A social strategy profile σ, together with D, determines a probability measure P_σ as follows.

For any $o = (u, p^1, p^2, ...)$, let

$$o^1 = (u),$$

and

$$o^t = (u, p^1, p^2, ..., p^{t-1}).$$

The Borel field over social type profiles induces a σ-algebra, F^1, over the space of all outcomes. Similarly, given the finite set of play paths at each time, we have an obvious σ-algebra, F^t, over outcomes based on the social type profile and the information observable through the beginning of stage t, for each $t > 1$. The σ-algebra on the space of all outcomes, F^∞, is the one generated by the sequence F^t. To define the probability measure P_σ over the space of all outcomes, we define P_σ^t for events in F^t for each t, inductively.

For any $A \in F^1$ let

$$P_\sigma^1(A) = D(\{u : u = o^0, o \in A\})$$

and, inductively, for any $A \in F^{t}$ let

$$P^t_\sigma(A) = \int_{o \in A} g_{\sigma(o^t)}(p^t_o)\, dP^{t-1}_\sigma(o),$$

where p^t_o is the play path at stage t under outcome o, and $\sigma(o^{t})$ is profile of behavioral strategies $(\sigma_{(1,h^t)}(u_{1,h^t}), \ldots, \sigma_{(n,h^t)}(u_{n,h^t}))$, where h^t and u_{i,h^t} correspond to the outcome o^{t}.

P_σ is the extension of the sequence P^{t}_{σ} to the field F^∞.

Equilibrium in the Bayesian Recurring Game

An *equilibrium* of the Bayesian recurring game is a profile of social strategies σ such that $\sigma_{i,h^t}(u_{i,h^t})$ maximizes the expected utility (u_{i,h^t} conditional on h^t) for each u_{i,h^t}. That is, for each (i, h^{t})

$$\begin{aligned} &\int_{u_{-i,h^t}} \left[\sum_{p^t} g_{\sigma_{-i,h^t}(u_{-i,h^t}), \sigma_{i,h^t}(u_{i,h^t})}(p^t) u_{i,h^t}(p^t) \right] dP_\sigma(u_{-i,h^t} | h^t) \\ \geq\ &\int_{u_{-i,h^t}} \left[\sum_{p^t} g_{\sigma_{-i,h^t}(u_{-i,h^t}), s_i}(p^t) u_{i,h^t}(p^t) \right] dP_\sigma(u_{-i,h^t} | h^t) \end{aligned}$$

for all $s_i \in S_i$.

3 Convergence to Equilibrium: A Purification Theorem

Harsanyi (1973) introduced the idea of a purification of a mixed strategy equilibrium, where a player may be any one of an infinite number of variants whose privately known preferences are close to some publicly known preferences. In an equilibrium of this Bayesian game, the average play of these variant types, who each play pure strategies, is close to the mixed strategy Nash equilibrium of the complete information game relative to the publicly known preferences. The following definitions are the analogous concepts for the Bayesian recurring games.

$\varepsilon-$Variants

Fix $\varepsilon > 0$. For a type v_i, a vector of types v, a i-type profile u_i, and a social type profile u, let $v^{\varepsilon}_{i}, v^\varepsilon, u^\varepsilon_i$, and u^ε denote the corresponding sets of

ε-variants, i.e., $v_i^{\varepsilon} = \{v_i' : \left|v_i' - v_i\right| \leq \varepsilon\}$, $v^{\varepsilon} = \{v' : \left|v' - v\right| \leq \varepsilon\}$, $u_i^{\varepsilon} = \{u_i' : \left|u_i' - u_i\right| \leq \varepsilon\}$ and $u^{\varepsilon} = \{u' : \left|u' - u\right| \leq \varepsilon\}$.

Given $v \in V$, a vector of stage game types, let G_v denote the complete information stage game defined by these types. Also, let $\vec{v}$ denote the constant social type profile defined by v (i.e., $\vec{v}_{i,h^t} \equiv v_i$ for every i and h^t) and σ be a fixed social strategy profile.

Playing Like

Given $\vec{v}^{\varepsilon}$, let $P_{\sigma|\vec{v}^{\varepsilon}}(A) \equiv P_{\sigma}(A|u \in \vec{v}^{\varepsilon})$ for every event A, and for any h^t let $P_{\sigma|\vec{v}^{\varepsilon},h^t}(p) \equiv P_{\sigma|\vec{v}^{\varepsilon}}(p|h^t)$ for every stage game play path p.

Fix $\gamma > 0$ and two distributions over stage game play paths $\tilde{g}$ and g. We say that $\tilde{g}$ *plays* γ*-like* g if $\max_p |\tilde{g}(p) - g(p)| \leq \gamma$.

Given $\gamma > 0$, $P_{\sigma|\vec{v}^{\varepsilon}}$ *eventually plays* γ*-like a Nash equilibrium of* G_v if for $P_{\sigma|\vec{v}^{\varepsilon}}-$a.e. outcome there exists T such that for each $t \geq T$ $P_{\sigma|\vec{v}^{\varepsilon},h^t}$ plays γ-like g_s for some Nash equilibrium s of G_v.

Theorem 1. *Consider an equilibrium of the Bayesian recurring game, σ, and a vector of stage game types v such that $\vec{v}$ is non-isolated. For every $\gamma > 0$ there exists $\varepsilon^* > 0$, such that if $0 < \varepsilon \leq \varepsilon^*$, then $P_{\sigma|\vec{v}^{\varepsilon}}$ eventually plays $\gamma-$like a Nash equilibrium of G_v.*

Remark 1: On the prior distribution. The prior probability distribution over the social type profile may be of a general form. It allows independent type selections from one period to the next and even the possibility that these independent selections use different distributions. Moreover, since a convex combination of priors is itself a prior, the model allows for uncertainty about "how society was selected." For example, in the prisoners' dilemma game we can write a prior describing the beliefs that, with probability ρ, all the future players will use the standard prisoners' dilemma table to assess their payoffs, and with probability $1-\rho$ every player born is equally likely (independently of past players) to be a tit-for-tatter or a standard prisoners' dilemma type. The non-isolation assumption in the statement of the theorem, that the selected type has positive probability neighborhoods, links periods in a way that makes learning possible.

Remark 2: On learning and purification: Considering play of the $\varepsilon-$variants of $\vec{v}$ rather than the play of $\vec{v}$ is critical to the validity of Theorem 1.

Jordan (1993) makes it clear that it will be impossible to have players learn to play mixed strategy equilibria, since they generally will have pure best responses to their beliefs at any stage. A recent paper by Nachbar (1996) (see also Foster and Young (1996)) builds on somewhat similar insight to point out difficulties in obtaining very strong learning conclusions: If one admits a sufficiently (uncountably) rich set of types, then one cannot conclude that players' will make arbitrarily correct predictions about the actions of the player that they face and best respond to those predictions. Here, we have relaxed the strong requirement that players make correct predictions about the actual player they face, and instead find that play averaged over types in any arbitrarily small neighborhood of a given type converges to Nash equilibrium play. [6]

Remark 3: A Corollary on Expectations: Since P_{σ} merges[7] with $P_{\sigma|\vec{v}^{\varepsilon}}$, we can modify the statement of Theorem 1 to conclude that for $P_{\sigma|\vec{v}^{\varepsilon}}$−a.e. outcome there exists T such that for each $t \geq T$ $P_{\sigma|h^t}$ plays γ-like g_s for some Nash equilibrium s of G_v. Then, in the case where almost every social profile is non-isolated, we can conclude that P_{σ} eventually plays γ-like a Nash equilibrium of G_v (where v is the realized stage utility profile). This provides a result similar to that of Jordan (1991), but for the case where the stage game is itself an extensive form game and only its play path is publicly observed.

4 Proof of Theorem 1.

A direct proof of Theorem 1 is possible but complicated. It involves combining a delicate backward induction type argument (within the stage games) with repeated use of Bayes' formula and the law of total probability, taking limits as ε goes to zero and T goes to infinity. However, the purification idea, combined with the concept of $\delta-\varepsilon-$subjective equilibrium, and a result on the merging of measures provides a natural, or even "simple," proof.

Fix v to be a vector of types for the stage game G_v. Given two strategy profiles s and r for $G(v)$, we say that *s plays γ-like r* if g_s plays γ-like g_r.

[6] See Nyarko (1994) for a different approach where beliefs are not necessarily correct in any single period, but match the empirical distribution of play over time.

[7] For $P_{\sigma|\vec{v}^{\varepsilon}}$−a.e. outcome and any $\gamma > 0$ there exists T such that for $t \geq T$, $P_{\sigma|h^t}$ plays γ−like $P_{\sigma|\vec{v}^{\varepsilon},h^t}$.

For each player role i consider a profile $s^i = (s_1^i, \ldots, s_n^i)$ of strategies for the stage game and let $s = (s_1^1, s_2^2, ..., s_n^n)$. s_j^i is interpreted as the strategy that player i believes player role j plays, and s, under the assumption that players know their own strategies, is the vector of strategies actually used. The following concept is defined by Kalai and Lehrer (1993b): [8]

A profile $(s_j^i)_{i,j}$ forms a $\delta - \varepsilon-$*subjective equilibrium* of the stage game, G_v, if for each i

(i) s^i plays δ-like s, and

(ii) s_i^i is an ε–best response to s_{-i}^i.

The following Lemma strengthens a claim (Remark 2) of Kalai and Lehrer (1993b), and is useful in the proof of Theorem 1.

Lemma 1. *Consider a stage game and a utility vector v. For every $\gamma > 0$ there exists $\delta^* > 0$ and $\varepsilon^* > 0$ such that each $\delta - \varepsilon-$subjective equilibrium with $\delta \leq \delta^*$ and $\varepsilon \leq \varepsilon^*$ plays γ-like some Nash equilibrium of the stage game G_v.*

Proof: Pick any $\gamma > 0$. By Proposition 2 in Kalai and Lehrer (1993b) there exists $\delta^* > 0$ such that each $\delta - 0-$ subjective equilibrium with $\delta \leq \delta^*$ plays γ-like some Nash equilibrium of the stage game G_v. Suppose that Lemma 1 is false. Then, there exist sequences $\delta^k \to 0$, $\epsilon^k \to 0$, and a corresponding sequence $\{(s_j^i)_{i,j}^k\}_k$ of $\delta^k - \varepsilon^k-$subjective equilibria that do not play γ-like any Nash equilibrium of G_v. Given the finiteness of the stage game, find a convergent subsequence of $\{(s_j^i)_{i,j}^k\}_k$ and let its limit be $(\bar{s}_j^i)_{i,j}$. It follows that $(\bar{s}_j^i)_{i,j}$ does not play γ-like any Nash equilibrium of G_v. It also follows that $(\bar{s}_j^i)_{i,j}$ is a $\delta^* - 0-$subjective equilibrium (since payoffs are continuous in mixed strategies). This contradicts the fact that any $\delta^* - 0-$ subjective equilibrium plays γ-like some Nash equilibrium of the stage game G_v. ∎

The proof of Theorem 1 is then based on the following construction. Consider the (non-Bayesian) recurring game, $R_{\vec{v}}$, with stage games G_v. For each player role i consider the ε^*-variants of v_i. Construct an i-player mixed strategy profile η_i in $R_{\vec{v}}$ which corresponds to the average play of the ε^*-variants of v_i in the Bayesian recurring equilibrium. The play generated by

[8] Also, see Battigali (1987) and Fudenberg and Levine (1993) for the related concepts of conjectural and self-conforming equilibria.

the constructed profile of strategies η induces the same distribution over the play paths of all stage games as does the Bayesian recurring game equilibrium when restricted to the ε^*-variants of v. Thus, it suffices to show that these constructed strategies η play γ-like a Nash equilibrium of $R_{\vec{v}}$ in all late stage games, which is the use of the purification idea. Relying upon Lemma 1, this is accomplished by starting with a small enough ε, and then showing that in late subgames the constructed strategies η are a δ-ε-subjective equilibrium for small enough δ. The δ-convergence of beliefs follows from results on merging.

Formally, for every measurable set of i-type profiles B_i define the B_i average strategy (induced by σ) $\eta_{i,h^t|B_i}$ by

$$\eta_{i,h^t|B_i} = \int_{u_{i,h^t}} \sigma_{i,h^t}(u_{i,h^t}) dP_\sigma(u_{i,h^t}|h^t, u_i \in B_i).$$

Lemma 2: *Fix $\varepsilon > 0$, $\delta > 0$, and a vector of non-isolated types $\vec{v}$. For almost every history generated by the strategies $(\eta_{i,h^t|v_i^\varepsilon})_i$ there is a time T such that for all $t \geq T$ the matrix defined below is a δ-ε-subjective equilibrium of G_v:*

$$s_j^i = \eta_{j,h^t|U_j} \text{ if } j \neq i \text{ and } s_i^i = \eta_{i,h^t|v_i^\varepsilon}.$$

The probability distribution over play paths induced by the strategies $s_i^i = \eta_{i,h^t|v_i^\varepsilon}$, is the distribution generated by the $\varepsilon-$variants of $\vec{v}$ under the Bayesian equilibrium. The distribution generated by the strategies $s_j^i = \eta_{j,h^t|U_j}$ is the distribution generated on (conditionally weighted) average by all types of j under the Bayesian equilibrium, and corresponds to the beliefs of i about the other players.

Proof of Lemma 2: Part (ii) in the definition of δ-ε-subjective equilibrium is satisfied by $(s_j^i)_{i,j}$ by the definition of $\eta_{i,h^t|v_i^\varepsilon}$. (Each action in the support of $\eta_{i,h^t|v_i^\varepsilon}$ is a best response to (s_{-i}^i) for some $\varepsilon-$variant of v_i, and is thus an ε-best response relative to v_i.) To see part (i) of the definition of δ-ε-subjective equilibrium, notice that each strategy $\eta_{i,h^t|U_i}$ is a mixed strategy composed of $\eta_{i,h^t|v_i^\varepsilon}$ and a second strategy ($\eta_{i|C}$ with C being the complement of v_i^ε) which assigns strictly positive probability to $\eta_{i,h^t|v_i^\varepsilon}$ under the non-isolation condition. A result about merging (e.g., Theorem 3 in Kalai and Lehrer (1993a)) is then sufficient to guarantee that Part (i) of the definition of δ-ε-subjective equilibrium is satisfied after a sufficiently long random time T. ∎

Proof of Theorem 1: Fix $\gamma > 0$ and v such that $D(\vec{v}^{\phi}) > 0$ for all $\phi > 0$.

By Lemma 1, find δ^* and ε^* such that for any $\delta \leq \delta^*$ and $\varepsilon \leq \varepsilon^*$ any $\delta-\varepsilon-$subjective equilibrium relative to v plays γ–like some Nash equilibrium of G_v. It follows from Lemma 2 applied to $\delta \leq \delta^*$ and $\varepsilon \leq \varepsilon^*$ that for $P_{\sigma|\vec{v}^{\varepsilon}}$ almost every outcome there exists T such that $P_{\sigma}(p^t \mid h^t, u \in \vec{v}^{\varepsilon})$ plays γ-like some Nash equilibrium of G_v for each $t \geq T$. ∎

5 Noncooperation in the Prisoners' Dilemma

Consider the situation where the stage game is a finitely repeated prisoners' dilemma. Following almost standard terminology, in each of the m-rounds each of the two players chooses one of two actions: C (cooperate) or F (fink).

Of special interest are the types whose utility is computed to be the sum of the payoffs of the m-rounds of their own stage game obtained from a standard prisoners' dilemma table. The following table describes such single round payoffs for player 1:

		Player 2: C	Player 2: F
Player 1	C	1	-1
	F	2	0

and the symmetrically opposed table describes the single round payoffs of a standard player-two type. Let $r_i, 1 = 1, 2$, denote such standard types and, as before, let $\vec{r}$ denote the corresponding social profile.

The *noncooperative play path* is one where both players play action F in each round of the prisoners' dilemma.

Theorem 2. *Consider D such that $\vec{r}$is non-isolated, and an equilibrium σ of the Bayesian recurring m-times repeated prisoners' dilemma game. There* *exists* $\varepsilon^* > 0$ *such that for every* $0 < \varepsilon \leq \varepsilon^*$

$$lim_{t \to \infty} P_{\sigma|\vec{r}^{\varepsilon}} \text{ (a noncooperative path at time } t) = 1.$$

Theorem 2 states that if player roles are really being filled with players who are close to having standard prisoners' dilemma payoffs, then play will converge to that of the Nash equilibrium in the sense that it will become arbitrarily likely that players will play noncooperatively. The stronger conclusion, that players will play noncooperatively with certainty after some time is not necessarily true. Consider, for instance a version of the KMRW world, where there are social types who are sometimes tit-for-tatters and sometimes rational players. In equilibrium rational players could mix on C (in the early rounds relative to m) with positive probability in every stage following every history. This follows along the KMRW reasoning. Players can never completely rule out tit-for-tat types after any finite history. If rational players did not mix on C, then a rational player who did play C would be believed with certainty to be a tit-for-tat player and could gain, contradicting equilibrium.

Notice that the conclusion of Theorem 2 is stronger than the conclusion of Theorem 1. In Theorem 1, ε^* depends on how close one would like to be to the Nash equilibrium play; i.e., ε^* depends on γ. In Theorem 2, we can fix a single ε^* and eventually play as close as one likes to the (unique) Nash equilibrium play path. This stronger conclusion is derived from the particular structure of the prisoners' dilemma. To prove Theorem 2, we introduce a new concept which is a strengthening of the concept of $\delta - \varepsilon-$subjective equilibrium.

Tight $\delta - \varepsilon-$Subjective Equilibrium

A profile $(s^i_j)_{i,j}$ forms a *tight* $\delta - \varepsilon-$*subjective equilibrium* of the stage game, G_v, if for each i

(i) s^i plays δ-like s, and

(ii) each pure strategy in the support of s^i_i is an $\varepsilon-$best response to s^i_{-i}.

The concept of *tight* $\delta - \varepsilon-$subjective equilibrium [9] places a restriction on the nature of the $\varepsilon-$best response: each strategy in the support of a player's

[9] The notion of "tightness" is useful in other contexts as well. See Jackson and Kalai (1995b) for the related notion of a tight ε–Bayesian equilibrium and its usefulness.

strategy must be a best response relative to some ε-variant of v. The usual definition of ε-best response allows for mixtures that place small probability on strategies that are far from being best responses. Such mixtures are ruled out under the tight definition. This stronger definition and the structure of the prisoners' dilemma allows us to strengthen Lemma 1 to find a uniform ε across γ:

Lemma 3. *Consider the prisoners' dilemma stage game with utility vector r. There exists $\varepsilon^* > 0$ such that for every $\gamma > 0$ there exists $\delta^* > 0$ such that each tight $\delta - \varepsilon -$subjective equilibrium with $\delta \leq \delta^*$ and $\varepsilon \leq \varepsilon^*$ plays $\gamma -$like the Nash equilibrium of the stage game G_r.*

Proof of Lemma 3: Suppose the contrary. Then there exists a sequence of $\varepsilon_k \to 0$, and a corresponding sequence of $\gamma_k > 0$, such that for every [10] $\delta > 0$ there exists a tight $\delta - \varepsilon_k -$subjective equilibrium that does not play $\gamma_k -$like the Nash equilibrium of G_r. Consider any k. Since the above statement holds for all $\delta > 0$, it follows that there exists a tight $0 - \varepsilon_k -$subjective equilibrium that does not play $\gamma_k -$like the Nash equilibrium of G_r. (Take the limit of a convergent subsequence of tight $\delta - \varepsilon_k -$subjective equilibria as $\delta \to 0$.)

Next, given the structure of the finitely repeated prisoners' dilemma, G_r, we can find ε^* such that if $|v - r| \leq \varepsilon^*$, then the following holds: If in a given round a player i expects finking with probability 1 in all subsequent rounds, then any best response to the player's beliefs relative to v_i results in the player finking in that round. (Finking in that round results in a higher payoff than cooperating in that round, and can do no worse in the continuation since the current continuation has finking with probability 1 in any remaining rounds.)

Finally, combining the two ideas above, choose k such that $\varepsilon_k < \varepsilon^*$ and consider a tight $0 - \varepsilon_k -$subjective equilibrium that does not play $\gamma_k -$like the Nash equilibrium of G_r. There exists a last round in which some player finks with probability less than 1 on the play path. Consider that round and some such player i. Given that i must expect finking with probability 1 in all subsequent rounds and that $\varepsilon_k < \varepsilon^*$, a best response for i to any v_i where $|v - r| \leq \varepsilon^*$ must be to fink in that round, which is a contradiction. ∎

To prove Theorem 2, we need to restate Lemma 2 for the concept of tight $\delta - \varepsilon -$subjective equilibrium.

[10] Note that any tight $\delta' - \varepsilon -$subjective equilibrium is a tight $\delta - \varepsilon -$subjective equilibrium, whenever $\delta' \leq \delta$.

Lemma 4: *Fix* $\varepsilon > 0$, $\delta > 0$ *and a vector of non-isolated types* $\vec{v}$. *For almost every history generated by the strategies* $(\eta_{i,h^t|v_i^\varepsilon})_i$ *there is a time* T *such that for all* $t \geq T$ *the matrix defined below is a tight* $\delta - \varepsilon-$*subjective equilibrium of* G_v:

$$s_j^i = \eta_{j,h^t|U_j} \text{ if } j \neq i \text{ and } s_i^i = \eta_{i,h^t|v_i^\varepsilon}.$$

Proof of Lemma 4: This follows the proof of Lemma 2, substituting the words "tight $\delta - \varepsilon-$subjective equilibrium" instead of "$\delta - \varepsilon-$subjective equilibrium." ▮

Proof of Theorem 2: Theorem 2 is proven by combining Lemma 3 with Lemma 4 in the same way that Lemma 1 and Lemma 2 are used to prove Theorem 1, except that Lemma 3 allows for a uniform choice of ε^* across $\gamma \to 0$. ▮

6 Concluding Remarks

Extensive form stage games present specific off-equilibrium path learning problems (see Fudenberg and Kreps (1988)). Here the recurring play of $m \times G$ leads to convergence to Nash equilibrium even though the off-the-equilibrium play is never observed. However, in order to obtain results concerning convergence to subgame perfect equilibrium play, stronger assumptions are needed since there are additional requirements on off-equilibrium-path play. Explicit trembling is used by Jackson and Kalai (1995a) to consider learning in Selten's chain store game and examine the reputation phenomena of Kreps and Wilson (1982) and Roberts and Wilson (1982). That analysis is in the context of a limited set of types, but it suggests that a similar approach may be possible in the context studied here.

The analysis in this paper, as well as that in much of the Bayesian learning literature, assumes that agents can observe the play in all previous periods. An interesting issue for future consideration is how results such as those presented here extend when agents' historical horizons are limited. [11] For instance, if players can only observe the last n play paths - but realize that

[11] This question was raised by Bob Aumann in a seminar presentation of this paper.

earlier players were reacting to earlier observations and use this additional information - will similar results hold?

References

Battigali, P. [1987], "Comportamento Razionale ed Equilibrio nei Giochi e nelle Situazioni Sociali," Unpublished Dissertation, Bocconi University, Milano.

Blackwell, D. and L. Dubins [1962], "Merging of Opinions with Increasing Information," *Annals of Mathematical Statistics*, Vol. 38, pp. 882–886.

Foster, D. and P. Young [1996], "Learning with Hazy Beliefs," WP-370, Johns Hopkins University.

Fudenberg, D. and D. Kreps [1988], "A Theory of Learning, Experimentation and Equilibrium in Games," mimeo: Stanford University.

Fudenberg, D. and D. Levine [1993], "Self–Confirming Equilibrium," *Econometrica,* Vol. 61, pp. 523–545.

Fudenberg, D. and E. Maskin [1986], "The Folk Theorem in Repeated Games with Discounting and Incomplete Information," *Econometrica,* Vol. 54, pp. 533-554.

Harsanyi, J.C. [1967–1968], "Games of Incomplete Information Played by Bayesian Players, Parts I–III," *Managment Science,* Vol. 14, pp. 159–182, 320–334, 486–502.

Harsanyi, J. C. [1973], "Games with Randomly Preturbed Payoffs: A New Rationale for Mixed Strategy Equilibrium Points," *Int. J. of Game Theory*, Vol. 2, pp. 1-23.

Jackson, M. and E. Kalai [1995a], "Recurring Bullies, Trembling, and Learning," in *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, ed. by W. Albers, P. Hammerstein, B. Moldovaner, and E. van Damme, Springer, pp. 171-184.

Jackson, M. and E. Kalai [1995b], "Social Learning in Recurring Games," CMSEMS Discussion Paper # 1138, Northwestern University, forthcoming in *Games and Economic Behavior*.

Jordan, J. [1991], "Bayesian Learning in Normal Form Games," *Games and Economic Behavior,* Vol. 3, pp. 60–81.

Jordan, J. [1993], "Three Problems in Learning Mixed–Strategy Equilibria," *Games and Economic Behavior,* Vol. 5, pp. 368–386.

Kalai, E. and E. Lehrer [1993a], "Rational Learning Leads to Nash Equilibrium," *Econometrica,* Vol. 61, pp. 1019–1045.

Kalai, E. and E. Lehrer [1993b], "Subjective Equilibrium in Repeated Games," *Econometrica,* Vol. 61, pp. 1231–1240.

Kreps, D., P. Milgrom, D.J. Roberts, and R. Wilson [1982], "Rational Cooperation in the Finitely–Repeated Prisoners' Dilemma," *Journal of Economic Theory,* Vol. 27, pp. 245–252.

Kreps, D. and R. Wilson [1982], "Reputation and Imperfect Information", *Journal of Economic Theory*, Vol. 27, 253-279.

Kuhn, H. [1953], "Extensive Games and the Problem of Information." In H. Kuhn and A.W. Tucker, *Contributions to the Theory of Games, Vol. II – Anals of Mathematics Studies 28*, Princeton University Press, pp. 193–216.

Lehrer, E. and R. Smorodinsky [1997], "Repeated Large Games with Incomplete Information," *Games and Economic Behavior,* Vol. 18, pp. 116–134.

Marimon, R. [1995], "Learning from Learning in Economics," forthcoming in *Proceedings of the 7th World Congress of the Econometric Society*.

Milgrom, P. and J. Roberts [1982], "Predation, Reputation, and Entry Deterrence", *Journal of Economic Theory*, Vol. 27, 280-312.

Nachbar, J. [1995], "Prediction, Optimization, and Rational Learning in Games," forthcoming in: *Econometrica.*

Sandroni, A. [1997], "The Almost Absolute Continuity Hypothesis," forthcoming in: *Games and Economic Behavior.*