Approximate Purification of Mixed Strategies
with Y. Katznelson, R. Radner, R. Rosenthal, and B. Weiss

1 Introduction

Randomized strategies play a significant role in the theory of games, but
have limited appeal in many practical situations. Practical situations,
however, often possess enough exogenous uncertainty to render explicit
randomization intuitively unnecessary. In this paper we bring these obser-
vations together by describing conditions under which randomized strate-
gies can be replaced by "approximately equivalent" pure strategies.

Suppose that \( n \) persons play a game in which each makes an obser-
vation (which may be related to the payoff) before play begins. While the
observations need not be independent, suppose that even after making
their observations and pooling the information so obtained, players
\( 2, \ldots, n \) still cannot ascribe positive probability to any particular possible
observation of Player 1. Then we will show that any mixed strategy of
Player 1 can be \( \varepsilon \)-purified, i.e., replaced by a pure strategy that yields all
players approximately the same payoff as the original mixed strategy, no
matter what strategies Players \( 2, \ldots, n \) use.

The two-person case of this result is proved in §4. The underlying idea
of the proof is described in §3. In §6 we present two alternative proofs,
which use Fourier analysis to construct \( \varepsilon \)-purifications somewhat more
explicitly. The general, \( n \)-person case follows easily from the two-person
case (§2).

Next, suppose only that from his own observation, no individual
player can ascribe positive probability to any particular observation of
any other player. Then in general, mixed strategies cannot be \( \varepsilon \)-purified
(§7). But we will show that every Nash equilibrium point in mixed stra-
tegies can be replaced by an \( n \)-tuple of pure strategies that is an approxi-
mate equilibrium (no player can gain much by deviating), and yields all
players approximately the same payoff as the original equilibrium point
(§5).

© The Operations Research Society of America and The Institute of Management Sciences,
290 Westminster Street, Providence, RI 02903. Reprinted with permission.
OR/MS Index 1978 subject classification. Primary 236 Games/group decisions/non-
cooperative.
Key words. Game theory, Nash equilibrium, pure strategies.
Research on this paper was begun during a visit of Rosenthal to the Institute for Advanced
Studies of the Hebrew University of Jerusalem in February 1980. Aumann’s research was
supported by the Institute for Advanced Studies, and by the Institute for Mathematical
Studies in the Social Sciences at Stanford University under National Science Foundation
Grants Nos. SOC 77-06000-A01 and SES 80-06654.
In §2 we spell out definitions and notations, and state our results precisely. §8 discusses the literature.

2 Definitions and Results

The two-player game is described as follows. Player 1 makes an observation from a set $X$, then selects an action from a finite set $K$. Player 2 makes an observation from a set $Y$, then selects an action from a finite set $L$. The set $X$ together with a $\sigma$-field $\mathcal{X}$ of subsets is a measurable space, assumed isomorphic to the unit interval of the real line together with its Borel $\sigma$-field. The same is true of $Y$ together with a $\sigma$-field $\mathcal{Y}$. A measure $\mu$ on the product $\sigma$-field $\mathcal{X} \otimes \mathcal{Y}$, called the prior, describes the common information of the players before the observations. Denote by $\mu_Y$ the marginal probability measure on $Y$ induced by $\mu$ (i.e., if $W \in \mathcal{Y}$ then $\mu_Y(W) = \mu(X \times W)$) and by $\mu(\cdot | y)$ any regular version of the conditional probability measure on $X$ given $y$. Call $\mu$ conditionally atomless for Player 1 if $\mu(\cdot | y)$ is atomless, $\mu_Y$-almost everywhere ($\mu_Y$-a.e.). The (generalized) payoff function is $u$, which takes $K \times L \times X \times Y$ into $n$-dimensional Euclidean space. We assume that $u(k, l, \cdot, \cdot)$ is $\mu$-integrable. (From now on we will usually omit the adjective "measurable" before the nouns "set," "function," and "partition." No confusion should result.)

Let $\Delta^K$ be the unit simplex in the finite-dimensional Euclidean space $E^K$, each coordinate of which corresponds to a distinct element of $K$. Then every point in $\Delta^K$ corresponds to a mixed action for Player 1. A strategy for Player 1 is a function $f$ from $X$ to $\Delta^K$. It is pure if its values are vertices of $\Delta^K$, $\mu_X$-a.e. The definitions for Player 2 are obvious analogues. If $f$ and $g$ are strategies for Players 1 and 2 respectively, the

1. Two measurable spaces are isomorphic if there is a one-to-one function between them that is measurable in both directions. This assumption enables us to avoid certain technical difficulties in defining conditional probability (see, e.g., Breiman [1968]). It is weaker than it looks; any Borel subset of any Euclidean space, and indeed of any complete separable metric space, satisfies it.

2. A measure is atomless if every set having positive measure contains a subset having strictly smaller but positive measure. If a measurable space is isomorphic to the unit interval with its Borel sets, then a measure is atomless if and only if every singleton has measure zero. It is this form of atomlessness which we use; see Lemma 1.

3. By presenting our basic result for vector-valued payoff functions, we are able to generate the $n$-person purification results as consequences of the basic result. The extra generality produces no extra difficulty in the proof. In an even more general setup the payoffs might depend on random elements not observed by any player. By taking $u$ to be the conditional expected payoff vector given all observations, our formulation is seen to be no less general in fact.

4. More traditionally, a "behavior strategy." Mixed and behavior strategies are equivalent in the context of this paper; see Aumann [1964].
expected payoff is
\[ U(f, g) = \int_{X \times Y} \sum_{k, l} f_k(x)g_l(y)u(k, l, x, y)\mu(dx \times dy). \]

Given \( \varepsilon > 0 \), two strategies \( f \) and \( f' \) are \( \varepsilon \)-equivalent for Player 1 if for all \( i \in \{1, \ldots, n\} \) and all strategies \( g \) of Player 2,
\[ |U_i(f, g) - U_i(f', g)| < \varepsilon, \]
where \( U_i \) denotes the \( i \)th component of \( U \), \( i = 1, \ldots, n \). An \( \varepsilon \)-purification of a strategy is a pure strategy \( \varepsilon \)-equivalent to it.

**Theorem** Suppose that \( \mu \) is conditionally atomless for Player 1. Then for every \( \varepsilon > 0 \), every strategy of Player 1 has an \( \varepsilon \)-purification.

The proof is found in §4.

We next extend the discussion to \( n \)-person games. There are \( n \) players \( (i = 1, \ldots, n) \); \( n \) observation spaces \( (X_i, \mathcal{X}_i) \), each isomorphic to the real unit interval with its Borel \( \sigma \)-field; and \( n \) action sets \( K_i \), each finite. The probability measure \( \mu \) is now defined on the product \( \sigma \)-field \( \mathcal{X} = \mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_n \). For each \( k \in K = K_1 \times \cdots \times K_n \), \( x \in X = X_1 \times \cdots \times X_n \), and \( i \in \{1, \ldots, n\} \), \( u_i(k, x) \) denotes the (scalar) payoff to Player \( i \) from the joint move \( k \) when the joint observation is \( x \). The joint payoff function is \( u = (u_1, \ldots, u_n) \). For each \( k \in K \), \( u(k, \cdot) \) is assumed \( \mu \)-integrable.

The mixed action set for \( i \) is \( \Delta^K \), the unit simplex in the Euclidean space indexed by the elements of \( K_i \). A strategy for Player \( i \) is a function \( f^i \) from \( X_i \) to \( \Delta^{K_i} \). It is pure if its values are vertices of \( \Delta^{K_i} \), \( \mu_{X_i} \)-a.e. The expected payoff to an \( n \)-tuple of strategies \( f = (f^1, \ldots, f^n) \) is
\[ U(f) = \int_X \sum_k f^1_k(x_1) \cdots f^n_k(x_n)u(k, x)\mu(dx), \]
where \( k = (k_1, \ldots, k_n) \) and \( x = (x_1, \ldots, x_n) \).

Given \( \varepsilon > 0 \), two strategies \( f^j \) and \( f'^j \) of player \( j \) are \( \varepsilon \)-equivalent if for each strategy \( (n - 1) \)-tuple \( g^{-j} \) of the players other than \( j \) and for each player \( i \in \{1, \ldots, n\} \),
\[ |U_i(f^j, g^{-j}) - U_i(f'^j, g^{-j})| < \varepsilon, \]
where \( (f^j, g^{-j}) \) denotes the strategy combination \( (g^1, \ldots, g^{j-1}, f^j, g^{j+1}, \ldots, g^n) \). As before, an \( \varepsilon \)-purification of a strategy is a pure strategy \( \varepsilon \)-equivalent to it.

The measure \( \mu \) is conditionally atomless for \( i \) if for (almost) every \( (n - 1) \)-tuple of observations \( y \) of the players other than \( i \), the conditional probability measure \( \mu(-|y) \) on \( X_i \) is atomless.
COROLLARY A Suppose that \( \mu \) is conditionally atomless for \( i \). Then for every \( \varepsilon > 0 \), every strategy of \( i \) has an \( \varepsilon \)-purification.

Proof Follows from the Theorem by taking \( X \) in the Theorem to be \( X_i \) here and \( Y \) in the Theorem to be the product of all \( X_j \) other than \( X_i \) here.

An \( \varepsilon \)-equilibrium point is an \( n \)-tuple \( f \) of strategies such that for each player \( i \) and each strategy \( g^j \) of \( i \),

\[
U_i(g^i.f^{-i}) \leq U_i(f) + \varepsilon.
\]

An equilibrium point is a 0-equilibrium point. An \( \varepsilon \)-purification of an equilibrium point \( f \) is an \( n \)-tuple \( f' \) of pure strategies such that every \( n \)-tuple \( f'' \) obtained from \( f \) by replacing some of the \( f^j \) by \( f'^j \) (including, of course, \( f' \) itself) is an \( \varepsilon \)-equilibrium point and satisfies

\[
|U_i(f'') - U_i(f)| < \varepsilon
\]

for all \( i \). In words, any group of players can switch to pure strategies without appreciably affecting the payoff to anybody and while maintaining approximate equilibrium.

Let \( \mu_{ij} \) denote the marginal probability measure on \( X_i \times X_j \) induced by \( \mu \). The measure \( \mu \) is weakly conditionally atomless for \( i \) if for every \( j \) other than \( i \) and (almost) every observation \( y_j \) of \( j \), the conditional probability measure \( \mu_{ij}(.|y_j) \) on \( X_i \) is atomless. In words, after making his own observation, every player other than \( i \) perceives an atomless distribution for \( i \)'s observations.

COROLLARY B Suppose that \( \mu \) is weakly conditionally atomless for all players. Then for every \( \varepsilon > 0 \), every equilibrium point has an \( \varepsilon \)-purification.

Corollary B is proved in §5. In §7 we show that weak conditional atomlessness is not sufficient for the \( \varepsilon \)-purification of strategies.

3 Discussion

Consider the case in which there are just two players, with two possible actions each. Take the observation sets \( X \) and \( Y \) of both players to be the unit interval \([0, 1]\), and suppose the payoffs are independent of the observations, but depend nontrivially on Player 1's action. Even in this simple case it is not always possible to purify mixed strategies exactly. For a counterexample, taken from Radner and Rosenthal [1982], let \( \mu \) be twice Lebesgue measure on the right triangle above the diagonal, i.e., \( \{(x,y) : y \geq x\} \), and 0 elsewhere. Then \( \mu(.|y) \) is \( 1/y \) times Lebesgue measure on the interval \([0,y]\), and 0 elsewhere. If it were possible for Player 1
to purify (exactly) the mixed strategy in which he plays 1/2−1/2 independent of his observation, then there would exist a subset $S$ of $X$ with
\[ \mu(S|y) = 1/2 \]
for almost all $y$, i.e., a set that intersects each interval in exactly half its Lebesgue measure; and it is known that there is no such set.

Now while Player 1 cannot purify exactly, he can purify approximately. Indeed, setting
\[ S^m = \left[ 0, \frac{1}{2m} \right] \cup \left[ \frac{2}{2m}, \frac{3}{2m} \right] \cup \ldots \cup \left[ \frac{2m-2}{2m}, \frac{2m-1}{2m} \right] \]
(the set consisting of alternate intervals of length $1/2m$), we find
\[ \mu(S^m|y) \to \frac{1}{2} \quad (\text{as } m \to \infty) \]
for almost every $y$. This implies that for each $\varepsilon > 0$, for sufficiently large $m$ the pure strategy corresponding in the obvious way to $S^m$ is an $\varepsilon$-purification of the 1/2−1/2 strategy. The same is true whenever the measures $\mu(\cdot|y)$ are absolutely continuous, i.e., whenever $\mu$ is absolutely continuous with respect to Lebesgue measure on the square $X \times Y$.

Unfortunately, without the absolute continuity it need not be true that, a.e., $\mu(S^m|y) \to 1/2$. For example, let $T$ be the set of all $x$ in $X$ with a dyadic representation $\sum_{i=1}^{\infty} x_i/2^i$ in which $x_i = 1$ whenever $i$ is odd. Then $T$ is in a natural one-one correspondence with $[0, 1]$, and so there is an atomless measure on $T$ corresponding to Lebesgue measure on $[0, 1]$; equivalently, there is an atomless measure $\zeta$ on $X$ with support $T$. Define $\mu$ on $X \times Y$ by $\mu(\cdot|y) = \zeta$ for all $y$. Now $T \cap S^{2^j}$ is empty for even $j$, hence $\zeta(S^{2^j}) = 0$, and hence $\mu(S^m|y)$ cannot converge to 1/2.

In this particular case the proof can be rescued by using the fact that $\zeta$, like any atomless measure on $[0, 1]$, transforms into Lebesgue measure under an appropriate measurable automorphism of $[0, 1]$. But it is possible to choose $\mu$ so that all the $\mu(\cdot|y)$ are mutually singular to each other and to Lebesgue measure, or have mutually singular components; and then no single transformation can simultaneously transform them all to absolutely continuous measures.\(^5\)

To overcome this problem we proceed, as before, to cut up Player 1's observation space $X$ into $2m$ intervals of equal length; but rather than choosing alternate intervals with certainty, we choose each interval with

\(^5\) Let $\mu(\cdot|y)$ be the measure on $[0, 1]$ induced by Bernoulli trials with probability $y$, i.e., the measure of a dyadic interval $[j/2^{n},(j+1)/2^{n}]$ is $y^{i}(1-y)^{n-i}$ where $i$ is the number of 1's in the dyadic representation of $j$. When $y = 1/2$ this is Lebesgue measure, but otherwise it is singular to Lebesgue measure; moreover, all the $\mu(\cdot|y)$ are mutually singular. Note that $\mu(S^{2^j}|y) = y$ for all $j$, so that $\mu(S^m|y)$ almost never converges to 1/2.
probability 1/2, independently. Denoting the random union of intervals so chosen by \( S_m \), we conclude as in the law of large numbers that for almost all \( y \)—indeed, whenever \( \mu(\cdot | y) \) is atomless—almost surely \( \mu(S_m | y) \to 1/2 \).

By Fubini’s theorem we can reverse these “almost universal” quantifiers and conclude that the probability is 1 that for almost all \( y \), \( \mu(S_m | y) \to 1/2 \). But since this has probability 1, it holds for at least one specific realization \( \{S_1, S_2, \ldots \} \) of the random sequence \( \{S_1, S_2, \ldots \} \). That is, \( \mu(S_m | y) \to 1/2 \) for almost all \( y \), which is what we need.

Note that we have not actually constructed a “purifying sequence” \( \{S_1, S_2, \ldots \} \)—one such that \( \mu(S_m | y) \to 1/2 \) for almost all \( y \). In §6 we present two alternative approaches. First, we show that the sequence \( \{S^m\} \), whose members consist of alternate intervals of length \( 1/2m \), has a purifying subsequence; in fact, one with asymptotic density 1. Second, multiplying each \( S^m \) by the same positive constant \( \beta \) (i.e., considering alternate intervals of length \( \beta/2m \) rather than \( 1/2m \)), we then use the same basic idea as above (Fubini’s theorem) to show that \( \{\beta S^m\} \) is itself purifying for almost all \( \beta \). While these “constructions” are not entirely explicit either, they do give us a better idea of how \( \varepsilon \)-purifications might look.\(^6\)

For a “practical” example in which \( \mu \) does not satisfy the absolute continuity condition, consider a 3-person game in which Player 1 observes \( \gamma_2 \) and \( \gamma_3 \), Player 2 observes \( \gamma_1 \) and \( \gamma_3 \), and Player 3 observes \( \gamma_1 \) and \( \gamma_2 \), each \( \gamma_i \) being uniformly and independently distributed on \([0,1]\). Since \( \mu \) is supported on the possible triples of observations and these form a 3-dimensional subset of the 6-dimensional product of the observation spaces, it cannot be absolutely continuous. Here \( \mu \) is not conditionally atomless for any player, but it is weakly conditionally atomless for all players; hence Corollary B applies, and thus all equilibria can be \( \varepsilon \)-purified.

---

6. Our methods do enable the actual construction, in the sense of Turing, of a purifying sequence. For example, since \( \{S^m\} \) has a purifying subsequence, for each \( k \) there is an \( m_k \) such that \( |\mu(S^m | y) - 1/2| < 1/k \) except possibly for a set of \( y \) of measure \( 1/2^k \) at most. Assuming that the \( \mu(S^m | y) \) are explicitly given and calculable in the appropriate sense, we can just try one \( m \) after another until we reach an appropriate \( m_k \). Similarly the first proof can be adapted to construct a purifying sequence, since each \( S_m \) has only finitely many realizations.

Note that a sequence can only be purifying “relative to \( \mu \)” (i.e., to the family \( \{\mu(\cdot | y)\} \)); there can be no “absolutely purifying” sequence \( \{S_m\} \), i.e., one such that \( \zeta(S_m) \to 1/2 \) for all atomless probability measures \( \zeta \). This follows from a theorem of Erdos, Kestelman, and Rogers [1963], according to which every sequence \( \{S_m\} \) of sets (in the unit interval) whose Lebesgue measure \( \lambda(S_m) \) is bounded away from zero has a subsequence whose intersection contains a perfect set \( P \). An absolutely purifying sequence \( \{S_m\} \) would have to satisfy the hypothesis, since \( \lambda(S_m) \to 1/2 \) (if necessary we could discard a finite number of \( S_m \)). Hence the conclusion is satisfied; if, then, \( \zeta \) is an atomless probability measure with support \( P \), then \( \zeta(S_m) \to 1 \) for infinitely many \( m \), and so \( \zeta(S_m) \to 1/2 \).
4 Proof of the Theorem

Without loss of generality (w.l.o.g.) we take \((X, \mathcal{X})\) to be the unit interval with its Borel sets. In the lemmas, \(\nu\) is an arbitrary but fixed probability measure on \(X \times Y\), with the property that \(\nu(\cdot|y)\) is atomless, \(\nu_Y\)-a.e.; and throughout this section, \(\varepsilon\) is an arbitrary but fixed positive real number.

**Lemma 1** Let \(T \in \mathcal{X}\). Then there exists a partition \(H_M = \{H_M^1, \ldots, H_M^M\}\) of \(T\) such that

\[
\int_Y \left( \max_{1 \leq j \leq M} \nu(H_M^j|y) \right) \nu_Y(dy) < \varepsilon^2.
\]

**Proof** For \(m = 1, 2, \ldots\), let the partition \(H_m\) be composed of the elements

\[
H_m^1 = T \cap \left[0, \frac{1}{m}\right], \quad H_m^2 = T \cap \left[\frac{1}{m}, \frac{2}{m}\right], \ldots, \quad H_m^m = T \cap \left[1 - \frac{1}{m}, 1\right].
\]

Since \(\nu(\cdot|y)\) is atomless, the cumulative distribution function \(\nu([0,x]|y)\) is continuous, and so uniformly continuous, in \(x\). Hence

\[
\lim_{m \to \infty} \left( \max_{1 \leq j \leq m} \nu(H_m^j|y) \right) = 0.
\]

From Lebesgue’s dominated convergence theorem,

\[
\lim_{m \to \infty} \int_Y \left( \max_{1 \leq j \leq m} \nu(H_m^j|y) \right) \nu_Y(dy) = 0,
\]

and the required inequality follows.

Denote the ordinary Euclidean norm by \(\|\|\).

**Lemma 2** Let \(s\) be any vector in \(\Delta^K\). Then for every \(T \in \mathcal{X}\) there is a function \(b\) from \(T\) to the vertices \(V\) of \(\Delta^K\) such that

\[
\int_Y \left\| s - \int_T b(x)\nu(dx|y) \right\| \nu_Y(dy) < \varepsilon.
\]

**Proof** Let \(H_m\) be as in the proof of Lemma 1 for \(m = 1, 2, \ldots\). Consider an auxiliary probability space \((\Omega, \mathcal{F}, P)\) and a sequence \(Z^1, Z^2, \ldots\) of independent random variables on this space taking values in \(V\) with

\[
E(Z^j) = s \quad \text{for } j = 1, 2, \ldots
\]
(where $E$ denotes expectation with respect to $P$). Next, define a function $b_m$ on $T$ by

$$b_m(x) = Z^j \quad \text{if} \quad x \in H^j_m.$$ 

Thus $b_m$ takes values in a space of random variables on $(\Omega, \mathcal{F})$. For every $y$,

$$\int_T b_m(x) v(dx|y) = \sum_{j=1}^m v(H^j_m|y)Z^j.$$ 

Hence, for all $y$

$$E\left(\int_T b_m(x) v(dx|y)\right) = s.$$ 

Similarly, for all $y$ (and denoting by $Z^j_k$ the $k$th coordinate of $Z^j$),

$$E\left(\left\|s - \int_T b_m(x) v(dx|y)\right\|^2\right) = E\left(\left\|s - \sum_j v(H^j_m|y)Z^j\right\|^2\right)$$

$$= \sum_k \text{Var}\left[\sum_j v(H^j_m|y)Z^j_k\right]$$

$$= \sum_j v(H^j_m|y)^2 \sum_k \text{Var}(Z^j_k).$$

But, since $Z^j$ takes on values in $\Delta^K$, $\sum_k \text{Var}(Z^j_k) \leq 1$; hence

$$E\left(\left\|s - \int_T b_m(x) v(dx|y)\right\|^2\right) \leq \max_j v(H^j_m|y)$$

for every $y$, and so by Lemma 1 and Fubini's theorem,

$$\varepsilon^2 > \int_Y E\left(\left\|s - \int_T b_M(x) v(dx|y)\right\|^2\right) v_Y(dy)$$

$$= E \left[ \int_Y \left\|s - \int_T b_M(x) v(dx|y)\right\|^2 v_Y(dy) \right].$$

Hence there is a realization $b$ of $b_M$ such that

$$\varepsilon^2 > \int_Y \left\|s - \int_T b(x) v(dx|y)\right\|^2 v_Y(dy) > \left[ \int_Y \left\|s - \int_T b(x) v(dx|y)\right\| v_Y(dy) \right]^2,$$

by the Cauchy-Schwarz inequality. The desired conclusion follows.
Define a seminorm\(^7\) on the functions \(f\) from \(X\) to \(E^K\) by

\[
\|f\|_v = \int_Y \left\| \int_X f(x) v(dx|y) \right\| v_Y(dy).
\]

**Lemma 3**  For any strategy \(f\) of Player 1, there is a pure strategy \(f'\) with

\[
\|f - f'\|_v < \epsilon.
\]

**Proof**  If \(f\) has only one value \(s\), this is Lemma 2 with \(T = X\). If \(f\) is simple (i.e., takes on only finitely many values in \(\Delta^K\)), denote by \(Q\) the number of distinct values taken on by \(f\), replace \(\epsilon\) in Lemma 2 by \(\epsilon/Q\), and partition \(X\) into the \(Q\) sets over which \(f\) is constant. The pure strategy \(f'\) is then pieced together from the \(Q\) functions constructed in Lemma 2.

Finally, if \(f\) is any strategy, there is a simple strategy \(d\) such that

\[
\sup_X |f(x) - d(x)| < \epsilon,
\]

and hence \(\|f - d\|_v < \epsilon\). Thus in the \(\|\cdot\|_v\) seminorm, \(f\) is approximable by simple strategies. Since in the previous paragraph we showed that simple strategies are approximable by pure strategies, our conclusion follows.

**Proof of Theorem**  Assume first that \(u_i(k, l, x, y)\) is always nonnegative. If

\[
\sum_{k,l} \int_{X \times Y} u_i(k, l, x, y) \mu(dx \times dy) = 0,
\]

then all the \(U_i\) vanish identically, and there is nothing to prove. Otherwise, let \(Y'\) be the space \(Y \times \{1, \ldots, n\} \times K \times L\), with generic element \(y' = (y, i, k, l)\), and define a probability measure \(v\) on \(X \times Y'\) by

\[
v(dx \times dy') = \nu(dx \times dy \times \{(i, k, l)\}) = cu_i(k, l, x, y) \mu(dx \times dy),
\]

where the positive constant \(c\) is chosen to make \(v\) a probability measure. Since \(\mu\) is conditionally atomless for Player 1, so is \(v\). Replacing \(Y\) by \(Y'\) in Lemma 3 and in the definition of the seminorm \(\|\cdot\|_v\), we conclude that for every strategy \(f\) of Player 1 there is a pure strategy \(f'\) such that

\[
\|f - f'\|_v < \epsilon.
\]

Then for all strategies \(g\) of Player 2,

\[
\sum_{i=1}^n |U_i(f, g) - U_i(f', g)|
\]

\[
\leq \frac{1}{c} \int_Y g_i(y) \left[ \int_X |f_k(x) - f'_k(x)| \nu(dx|y') \right] v_Y(dy')
\]

\[
\leq \frac{1}{c} \|f - f'\|_v < \frac{\epsilon}{c},
\]

7. Functional with all the defining properties of a norm except \(\|x\| = 0 \Rightarrow x = 0\).
and the theorem follows. If \( u_i(k, l, x, y) \) has negative values, set \( u = u^+ - u^- \), where \( u^+ \) and \( u^- \) are nonnegative and integrable. Applying the "nonnegative theorem" just proven to the 2n-dimensional payoff \((u^+, u^-)\), we obtain the desired theorem for the original vector payoff \( u \).

5 Proof of Corollary B

**Lemma 4** Let \( G^1, \ldots, G^m \) be \( m \) different 2-person games, in each of which Player 1 has the same observation set \( X_0 \) and the same action set \( K_0 \). Suppose that each \( \mu^j \) is conditionally atomless for Player 1, where \( \mu^j \) is the probability measure on pairs of observations in the \( j \)th game \( G^j \). Then for every \( \epsilon > 0 \) and every strategy \( f \) of Player 1, there is a pure strategy \( f' \) that is \( \epsilon \)-equivalent to \( f \) simultaneously in all the games \( G^1, \ldots, G^m \).

**Proof** We may assume w.l.o.g. that in each \( G^j \) the payoff \( u^j \) is one-dimensional; for if it is \( n^j \)-dimensional, we can replace \( G_j \) by \( n^j \) games differing only in their payoffs.

Let \( G \) be the 2-person game in which chance first chooses one of the games \( G^j \) with probability \( 1/m \); Player 2, but not Player 1, is informed of chance's choice; and the chosen game is then played. The payoff in \( G \) is the \( m \)-dimensional vector whose \( j \)th component is 0 unless \( G^j \) was the game chosen by chance; and, in that case, the payoff in \( G \) is the payoff in \( G^j \).

Formally, denote the action and observation sets of Player 2 in \( G^j \) by \( L^j \) and \( Y^j \), respectively, and the payoff function by \( u^j \). W.l.o.g. let the \( Y^j \) be pairwise disjoint; and define \( Y = Y^1 \cup \cdots \cup Y^m \) and \( L = L^1 \times \cdots \times L^m \) (with generic element \( l = (l^1, \ldots, l^m) \)). In \( G \), the action sets for Players 1 and 2 are \( K_0 \) and \( L \), respectively; the observation sets are \( X_0 \) and \( Y \), respectively; the probability measure \( \mu \) on \( X_0 \times Y \) is given by

\[
\mu(dx \times dy) = \frac{1}{m} \mu^j(dx \times dy) \quad \text{when } dy \subset Y^j;
\]

and the payoff is the \( m \)-vector \( u \), the \( j \)th component of which is defined by

\[
u_j(k_0, l, x_0, y) = \begin{cases} u^j(k_0, l^j, x_0, y) & \text{when } y \in Y^j, \\ 0 & \text{otherwise}. \end{cases}
\]

Now let \( f \) be a strategy of Player 1, and let \( \epsilon > 0 \). By the Theorem, there is a pure strategy \( f' \) that is \( (\epsilon/m) \)-equivalent to \( f \) in \( G \). Let \( g^1, \ldots, g^m \) be any strategies for Player 2 in \( G^1, \ldots, G^m \), respectively. Define a strategy \( g \) for Player 2 in \( G \) by \( G(y) = g^j(y) \) if \( y \in Y^j \). Then for every \( j \),
\[ \frac{\varepsilon}{m} > |U_j(f, g) - U_j(f', g)| = \frac{1}{m} |U^j(f, g) - U^j(f', g')|, \]

so the proof of Lemma 4 is complete.

**Proof of Corollary B** The pure strategies \( f'^i \) are defined by an induction on the index \( i \). Suppose that \( f'^i \) has been defined for all \( i < m \), where \( 1 \leq m \leq n \). Let \( H^m \) be the set of all strategy \( n \)-tuples \( h \) that are obtained from \( f \) by replacing some subset (possibly empty) of the first \( m - 1 \) strategies \( f^i \) in the \( n \)-tuple \( f \) by the corresponding \( f'^i \). For each \( h \) in \( H^m \) and each \( j \) other than \( m \), define a 2-person game \( G^{mj} \) between \( m \) and \( j \) (in the roles of Players 1 and 2, respectively) by fixing the strategies of the players \( i \) other than \( m \) and \( j \) to be \( h^i \), and letting \( m \) and \( j \) play as in the original game. The payoff in \( G^{mj} \) is the \( n \)-dimensional vector resulting from the original game. Formally, in \( G^{mj} \) the observation sets of \( m \) and \( j \) are \( X_m \) and \( X_j \); the action sets are \( K_m \) and \( K_j \); the probability measure on \( X_m \times X_j \) is the marginal probability measure \( \mu_{mj} \) induced by \( \mu \); and the payoff \( u^{mj}(k_m, k_j, x_m, x_j) \) is the expectation of \( u(k, x) \) when the \( x_i \) other than \( x_m \) and \( x_j \) are jointly distributed according to the conditional probability measure \( \mu(\cdot|x_m, x_j) \) on \( \times_{i \neq m, j} X_i \), and the distribution of the \( k_i \) other than \( k_m \) and \( k_j \) is determined by the strategies \( h^i \). Applying Lemma 4 to the \((2n - m - 1)2^{m-2}\) games \( G^{mj} \) (remember that \( m \) is fixed in each single step of the induction), we deduce that there is a strategy \( f^m \) such that for all \( h \) in \( H^m \), all \( j \not= m \), all strategies \( g^j \) of \( j \), and all \( i \)

\[ |U_i^{mj}(f^m, g^j) - U_i^{mj}(f^m, g^j)| \leq \frac{\varepsilon}{2n-m+2}. \quad (5) \]

It is this \( f^m \) that is used for the \( m \)th step of the inductive definition of \( f' \).

Note that since \( f^m \) is now defined for all \( m \), including \( m = n \), the definition of \( H^m \) extends to \( m = n + 1 \) as well.

We now prove by induction on \( m \) that for \( m = 1, \ldots, n + 1 \), each \( h \) in \( H^m \) is an \((\varepsilon/2^{n-m+2})\)-equilibrium point and satisfies

\[ |U_i(h) - U_i(f)| < \frac{\varepsilon}{2n-m+2}. \quad (6) \]

For \( m = 1 \) this follows from the fact that in that case \( h = f \) and \( f \) is a 0-equilibrium point. Suppose it is true for \( m \) (with \( m \leq n \)); we will show it for \( m + 1 \). Suppose \( h' \in H^{m+1} \). If \( h^m = f^m \), then \( h' \in H^m \), and then our contention follows from the inductive hypothesis. Otherwise, \( h^m = f^m \). Let \( h \) in \( H^m \) be obtained from \( h' \) by replacing \( f^m \) by \( f^m \). Then for each \( j \not= m \) and each strategy of \( g^j \) of \( j \), we have

\[ U^{mj}(f^m, g^j) = U(g^j, h'^j), \]

\[ U^{mj}(f^m, g^j) = U(g^j, h'^j). \quad (7) \]
From the definition of $f^m$ and the fact that $h$ is an $(e/2^{n-m+2})$-equilibrium point, we deduce that $h'$ is an $(e/2^{n-(m+1)+2})$-equilibrium point; from (5), (7), and (6) we obtain

$$|U_i(h') - U_i(f)| < \frac{e}{2^{n-(m+1)+2}}$$

for all $i$; so our induction is complete.

The end of the induction—i.e., the case $m = n + 1$—asserts precisely that $f'$ is an $\epsilon$-purification of the equilibrium point $f$, so the proof of Corollary B is complete.

6 The Fourier Approach

In this section we indicate two alternative constructions of $\epsilon$-purifications of strategies that assign the same mixed action to each observation—the situation addressed in Lemma 2. From this, the proof of the Theorem can be completed as in §4. For simplicity and transparency, we confine ourselves in this section to the case in which Player 1 has just two actions, the strategy to be $\epsilon$-purified is his $1/2-1/2$ strategy, and $X = [0, 1]$.

As indicated in §3, constructing $\epsilon$-purifications of this strategy for arbitrarily small $\epsilon$ is equivalent to constructing purifying sequences of subsets of $X$, i.e., sequences $\{S_1, S_2, \ldots\}$ such that $\mu(S_m \mid y) \to 1/2$ for $\mu_y$-almost all $y$ (henceforth simply almost all $y$). Note that a sequence that is purifying for one $\mu$ may not be purifying for another $\mu$; and, indeed, our constructions will depend on $\mu$.

The starting point is again the sequence $\{S^1, S^2, \ldots\}$ defined in §3. If we subtract 1/2 from the indicator function of $S^m$, we get a function $f^m$ whose graph looks like a squared-off sine wave. Let $f$ be the periodic function on the real line that has period 1 and equals $f^1$ on $[0, 1)$; then $f^m(x) = f(mx)$ for $x \in [0, 1]$, and the sequence $\{S^m\}$ is purifying if and only if for almost all $y$,

$$\int_0^1 f(mx) \mu_y(dx) \to 0 \quad \text{as } m \to \infty,$$  \hspace{1cm} (8)

where $\mu_y = \mu(\cdot \mid y)$. If we replace $f$ in (8) by an (appropriately normalized) sine or cosine, the integral becomes a Fourier-Stieltjes (henceforth simply Fourier) coefficient of $\mu_y$, and thus (8) becomes the statement that the Fourier coefficients of the measures $\mu_y$ approach 0.

It has long been known that there are atomless measures whose Fourier coefficients do not tend to zero (Zygmund [1955, §§2.213, 5.714, 11.52]); and this jibes well with the discussion in §3, where we showed that there
are \( \mu \) for which \( \{S^m\} \) is not purifying. On the other hand, by a theorem of Wiener (Katznelson [1976, p. 42]), the squared absolute values of the Fourier coefficients of atomless measures \( \zeta \) do tend to 0 in Cesaro mean (i.C.m.), i.e.,

\[
\frac{1}{2M + 1} \sum_{-M}^{M} |\hat{\zeta}(m)|^2 \to 0 \quad \text{as } M \to \infty,
\]

where \( \hat{\zeta}(m) \) is the Fourier coefficient 8

\[
\hat{\zeta}(m) = \int_{0}^{1} e^{-2\pi imx} \zeta(dx).
\]

This implies that an appropriate subsequence of the \( \hat{\zeta}(m) \) tends to 0. Moreover, we can find such a subsequence that will work simultaneously for almost all the \( \mu_y \). Indeed, the \( |\hat{\mu}_y(m)| \) are bounded by 1; hence applying Lebesgue's dominated convergence theorem to (9), we find

\[
\frac{1}{2M + 1} \sum_{-M}^{M} \int_Y |\hat{\mu}_y(m)|^2 \mu_Y(dy) \to 0 \quad \text{as } M \to \infty.
\]

Hence for some sequence of integers \( \{m_j\} \),

\[
\int_Y [|\hat{\mu}_y(m_j)|^2 + |\hat{\mu}_y(-m_j)|^2] \mu_Y(dy) \to 0 \quad \text{as } j \to \infty.
\]

Again applying Lebesgue's theorem, we deduce that for almost all \( y \), \( |\hat{\mu}_y(m_j)| \to 0 \) and \( |\hat{\mu}_y(-m_j)| \to 0 \) as \( j \to \infty \), which in turn implies that for almost all \( y \), the \( m_j \)th sine and cosine coefficients of \( \mu_y \) tend to 0.

To finish our argument, it remains only to replace the true Fourier coefficients by the "squared-off" Fourier coefficients appearing in (8). We proceed by establishing Wiener's theorem for these squared-off coefficients. Let \( f^e \) and \( f^{-e} \) be the continuous periodic functions pictured in Figure 1. Let \( g^e \) and \( g^{-e} \) be trigonometric polynomials,\(^9\) also with period 1, that uniformly approximate \( f^e \) and \( f^{-e} \) respectively to within \( e \), and let \( a_0^\pm e \) be their constant terms. Integrating \( g^\pm e \) over \([0, 1]\) w.r.t. Lebesgue measure, we find \( |a_0^\pm e| \leq 2e \).

For a fixed atomless measure \( \zeta \) on \([0, 1]\), set

\[
c_m = \int_{0}^{1} f(mx)\zeta(dx),
\]

8. The complex form, from which the real sine and cosine coefficients are easily derived.
9. Finite sums of the form \( \sum_m a_m e^{2\pi imx} \). We use the Weierstrass approximation theorem (e.g., Katznelson [1976, p. 15]).
Figure 1

\( f \) has the solid graph, \( f' \) the dashed graph, and \( f^{-\varepsilon} \) the dotted graph. Where they are not seen, the dashed and dotted lines coincide with the solid line. The abscissas of some points are indicated; their ordinates are \( \pm \frac{1}{2} \).

\[
c_{m}^{\pm \varepsilon} = \int_{0}^{1} f^{\pm \varepsilon}(mx)\zeta(dx),
\]

\[
d_{m}^{\pm \varepsilon} = \int_{0}^{1} g^{\pm \varepsilon}(mx)\zeta(dx).
\]

The Cauchy-Schwarz inequality and (9) yield \( |\hat{\zeta}(m)| \to 0 \) i.C.m., hence \( |d_{m}^{\pm \varepsilon} - a_{0}^{\pm \varepsilon}| \to 0 \) i.C.m., and hence

\[
\limsup_{M \to \infty} \frac{1}{2M + 1} \sum_{-M}^{M} |c_{m}^{\pm \varepsilon}| \leq |a_{0}^{\pm \varepsilon}| + \varepsilon \leq 3\varepsilon. \tag{10}
\]

From \( f^{\varepsilon} \geq f \geq f^{-\varepsilon} \) we deduce \( c_{m}^{\varepsilon} \geq c_{m} \geq c_{m}^{-\varepsilon} \), and hence \( |c_{m}| \leq |c_{m}^{\varepsilon}| + |c_{m}^{-\varepsilon}| \). Hence by (10),

\[
\limsup_{M \to \infty} \frac{1}{2M + 1} \sum_{-M}^{M} |c_{m}| \leq 6\varepsilon;
\]

since the left side does not depend on \( \varepsilon \), it follows that \( |c_{m}| \to 0 \) i.C.m., which is the "squared-off" Wiener theorem we were seeking. Using Lebesgue's dominated convergence theorem as before, we deduce that (8) holds if \( m \to \infty \) through an appropriate sequence of integers, i.e.,

**Proposition 11**  For each \( \mu \), there is a purifying subsequence \( \{S^{m}_{j}\} \) of \( \{S^{m}\} \).

From the above proof one can actually get somewhat more. The subsequence \( \{m_{j}\} \) is not sparse; on the contrary, one can find a purifying subsequence with density 1. In other words, for given large \( m \) the chances are very good that \( S^{m} \) yields an \( \varepsilon \)-purifying strategy; the inappropriate,
"bad" $S^m$ are few and far between. This jibes well with the counterexample in §3, where the "bad" $m$ are the even powers of 2.

For the second construction we require an extension of the Wiener theorem in several directions. We have already noted that by the Cauchy-Schwarz inequality, $|\hat{\zeta}(m)|^2$ in (9) can be replaced by $|\hat{\zeta}(m)|$. Next, the interval of summation in (9) can be replaced by any interval of the same length, with a starting point that may depend arbitrarily on $M$; this is proved in the same way as (9). Third, Wiener's theorem may be extended to the continuous (Fourier transform) case (Katzenelson [1976, p. 138]). Putting all this together yields

$$\frac{1}{T} \int_{T_0}^{T_0+T} |\hat{\zeta}(\xi)| \, d\xi \to 0 \quad \text{as } T \to \infty, \text{ uniformly in } T_0,$$

for atomless measures $\zeta$, where $\hat{\zeta}(\xi)$ is the Fourier transform

$$\hat{\zeta}(\xi) = \int_0^1 e^{-i\xi x} \zeta(dx).$$

Proceeding as above, we can again replace $\hat{\zeta}(\xi)$ by the "squared-off" Fourier transform

$$c(\xi) = \int_0^1 f(\xi x) \zeta(dx),$$

and conclude that

$$\frac{1}{T} \int_{T_0}^{T_0+T} |c(\xi)| \, d\xi \to 0 \quad \text{as } T \to \infty, \text{ uniformly in } T_0. \quad (12)$$

For each $\alpha$ in $[1, 2]$, define $S^m_\alpha$ as the union of alternate intervals of length $1/2m\alpha$ in $[0, 1]$, starting with $(0, 1/2m\alpha)$; the last interval may be cut off in the middle.$^{10}$ Let $f^m_\alpha$ be the indicator function of $S^m_\alpha$, less $1/2$. Then

$$\int_0^1 f^m_\alpha(x) \mu_\alpha(dx) = \int_0^1 f(\alpha mx) \mu_\alpha(dx) = c_\alpha(\alpha m), \quad (13)$$

where

$$c_\alpha(\xi) = \int_0^1 f(\xi x) \mu_\alpha(dx).$$

Integrating (13) over $\alpha$ and using (12), we find

$10.$ $\beta S^m$ in §3 is $S^m_\alpha$ here, with $\alpha = 1/\beta$. 

\[ \int_1^2 \left| \int_0^1 f_x^m(x) \mu_y(dx) \right| d\alpha = \int_1^2 |c_y(\alpha m)| d\alpha = \frac{1}{m} \int_0^{2m} |c_y(\xi)| d\xi \to 0 \quad \text{as } m \to \infty. \]

Integrating over \( y \) and using Lebesgue's theorem, we find

\[ \int_0^1 \int_0^2 \left| \int_y^1 f_x^m(x) \mu_y(dx) \right| d\alpha \mu_y(dy) \to 0. \]

Using Fubini's theorem to interchange the integrations over \( \alpha \) and \( y \), we deduce that for almost all \( \alpha \), for almost all \( y \),

\[ \int_0^1 f_x^m(x) \mu_y(dx) \to 0. \]

The same proof works if \( \alpha \) varies over any interval with positive endpoints. Thus we have

**Proposition 14** For each \( \mu \), for almost all \( \alpha \), the sequence \( \{S^m_\alpha\} \) is purifying.

7 A Counterexample

In Corollary A, conditional atomlessness cannot be replaced by weak conditional atomlessness. Explicitly, we exhibit a 3-person game in which the prior measure \( \mu \) is weakly conditionally atomless for one of the players, but there is an \( \varepsilon > 0 \) and a strategy of that player that has no \( \varepsilon \)-purification. The game may easily be modified so that \( \mu \) is weakly conditionally atomless for all players and still the same conclusion holds.

In this game, Player 1 observes both \( y \) and \( z \), Player 2 observes only \( y \), and Player 3 observes only \( z \); here \((y, z)\) is distributed over the unit square in accordance with a probability measure \( \theta \) with almost all the conditional measures \( \theta(\cdot|y) \) and \( \theta(\cdot|z) \) atomless. The observation spaces \( Y \) and \( Z \) of Players 2 and 3 respectively both are the unit interval, whereas Player 1’s observation space \( Y \times Z \) is the unit square. The prior measure \( \mu \) is formally defined on the four-dimensional cube \((Y \times Z) \times Y \times Z\), but is actually supported on the two-dimensional “diagonal” of this cube. If we identify this diagonal with \( Y \times Z \), then when restricted to the diagonal, \( \mu \) becomes \( \theta \); thus the atomlessness of the \( \theta(\cdot|y) \) and \( \theta(\cdot|z) \) implies that \( \mu \) is weakly conditionally atomless for Player 1. On the other hand, \( \mu \) is not conditionally atomless for Player 1, since by pooling their information, Players 2 and 3 get to know the precise observation of Player 1.

Let all three players have the two-point action set \( \{0, 1\} \), and define the (one-dimensional) payoff to be 1 if all players choose the same action,
0 otherwise. Consider the strategy for Player 1 in which he chooses 1/2–1/2 no matter what his observation is. If this strategy can be approximately purified, then for every \( \varepsilon > 0 \) there exists a subset \( A \) of \( Y \times Z \) that comes within \( \varepsilon \) of cutting each rectangle\(^{11}\) in \( Y \times Z \) in half w.r.t. \( \theta \), i.e.,

for each \( B \subset Y \) and \( C \subset Z \), \[
|\theta(A \cap (B \times C)) - \frac{1}{2} \theta(B \times C)| < \varepsilon.
\]

(15)

We will construct a \( \theta \) for which this is false.

For given \( m \), let \( \mathcal{Q}_m \) denote the field of sets in the unit interval generated by the quartic intervals \( (i/4^m, (i+1)/4^m] \). A subset of the unit square \( Y \times Z \) is called quartic if, for some \( m \), it is in the product field \( \mathcal{Q}_m \). A subset of \( Y \times Z \) is called quartic if, for some \( m \), it is in the product field \( \mathcal{Q}_m \). (i.e., if it is a finite union of squares whose sides are quartic intervals). The quartic subsets of \( Y \times Z \) can be listed in a sequence \( \{A_0, A_1, A_2, \ldots\} \), such that \( A_m \in \mathcal{Q}_m \) for all \( m \).

We will define \( \theta \) gradually, starting with \( \mathcal{Q}_0 \) and extending the definition one step at a time, from \( \mathcal{Q}_m \) to \( \mathcal{Q}_{m+1} \), until it is defined for all quartic sets. Caratheodory's theorem is then used to extend \( \theta \) to all Borel sets. The idea is to construct the extension from \( \mathcal{Q}_m \) to \( \mathcal{Q}_{m+1} \) so that (15) fails for \( A = A_m \) when \( \varepsilon = 0.1 \). Thus when \( \theta \) is completely defined, (15) fails for all quartic \( A \) when \( \varepsilon = 0.1 \). Finally, an approximation is used to conclude that (15) fails for all \( A \) when \( \varepsilon = 0.05 \).

Specifically, start by defining \( \theta(Y \times Z) = 1 \); this defines \( \theta \) on \( \mathcal{Q}_0 \). Next, suppose \( \theta \) is defined on \( \mathcal{Q}_m \). Since \( A_m \in \mathcal{Q}_m \), each atom\(^{12}\) of \( \mathcal{Q}_m \) must be included either in \( A_m \) or in its complement \(-A_m \). Divide each such atom \( S \) into a \( 4 \times 4 \) checkerboard, with a black or white square in the bottom left corner according as \( S \subset A_m \) or \( S \subset -A_m \). Define \( \theta \) to be \( \theta(S)/8 \) on the resulting black squares, 0 on the white squares. This completes the inductive definition of \( \theta \) on all quartic sets.

Note that if \( B \) is in some \( \mathcal{Q}_l \), and \( Q \) is an atom of \( \mathcal{Q}_m \), then

\[
\theta(B \times Q) \leq 2^m \lambda(B) \lambda(Q) = \lambda(B) \sqrt[4]{\lambda(Q)},
\]

where \( \lambda \) is Lebesgue measure. To apply Caratheodory's theorem, let \( \{D_k\} \) be a nondecreasing sequence of quartic sets whose union is a quartic set \( D \); we must show that

\[
\theta(D_k) \to \theta(D).
\]

(17)

11. Rectangles in \( Y \times Z \) correspond to pairs of pure strategies of Players 2 and 3.
12. Nonempty member of \( \mathcal{Q}_m \) not containing any other nonempty member, i.e., a square of the form \( (i/4^m, (i+1)/4^m] \times (j/4^m, (j+1)/4^m] \).
W.l.o.g. let \( D = P \times Q \), where \( P \) and \( Q \) are atoms of a \( \mathcal{Q}_m \). Given \( \delta > 0 \), use (16)—and the corresponding inequality for \( \theta(P \times C) \)—to find a quartic neighborhood \( W \) of the boundary of \( D \) with \( \theta(W) < \delta \). Then

\[
D \setminus D_k \subset W
\]  

(18)

for sufficiently large \( k \); for if not, then the sets \( D \setminus D_k \) would have a point in common in the interior of \( D \), contrary to \( \bigcup_k D_k = D \). But (18) implies that \( \theta(D) - \theta(D_k) \leq \theta(W) \leq \delta \), and so proves (17). Thus \( \theta \) is defined for all \( A \subset Y \times Z \).

To show that almost all the conditionals \( \theta(\cdot|y) \) and \( \theta(\cdot|z) \) are atomless, use an approximation argument to show that (16) continues to hold when \( B \) is an arbitrary subset of \( Y \). Hence for each atom \( Q \) of any \( \mathcal{Q}_m \), for almost all \( y \) in \( Y \), we have \( \theta(Q|y) \leq \sqrt{\lambda(Q)} \). Since there are only denumerably many atoms \( Q \) in all the \( \mathcal{Q}_m \) put together, we may interchange quantifiers and conclude that for almost all \( y \), for all quartic intervals \( Q \) (i.e., atoms of any \( \mathcal{Q}_m \)), we have \( \theta(Q|y) \leq \sqrt{\lambda(Q)} \). But this implies that \( \theta(\cdot|y) \) is atomless, and similarly almost all the \( \theta(\cdot|z) \) are atomless. Of course, it follows that \( \theta \) itself is atomless, though this can also be shown directly.

Suppose that one of the \( A_m \) satisfies (15) with \( \epsilon = 0.1 \). By taking \( B = Y \) and \( C = Z \), we find \( \theta(A_m) > 0.4 \). Next, take both \( B \) and \( C \) to consist of alternate intervals of length \( 1/4^{m+1} \), starting with the second and ending with the last such interval. By the construction of \( \theta \), we have, for each atom \( S \) of \( \mathcal{Q}_m^2 \),

\[
\theta(S \cap (B \times C)) = \begin{cases} 
\theta(S)/2, & \text{if } S \subset A_m, \\
0, & \text{if } S \notin A_m.
\end{cases}
\]

Hence

\[
\theta(A_m \cap (B \times C)) = \frac{1}{2} \theta(A_m),
\]

\[
\theta((-A_m) \cap (B \times C)) = 0,
\]

and so

\[
\theta(A_m \cap (B \times C)) - \frac{1}{2} \theta(B \times C) = \frac{1}{2} \theta(A_m \cap (B \times C)) > 0.1,
\]

contradicting (15) with \( \epsilon = 0.1 \).

Finally, if \( A \) is an arbitrary Borel set, there is a quartic set \( A_m \) with \( \theta(A_m \Delta A) < 0.05 \), where \( \Delta \) denotes the symmetric difference. Then (15) fails for \( A \) when \( \epsilon = 0.05 \).

To construct an example in which \( \mu \) is weakly conditionally atomless for all players, let \( w \) be a random variable that is uniformly distributed over \([0, 1]\), is independent of \((y, z)\) and is observed by Players 2 and 3 but
not by Player 1; the action sets and payoffs remain as before (\(w\) does not affect the payoffs). This is similar to the situation in the last paragraph of §3. In particular, the prior measure is weakly conditionally atomless for all players; but the strategy of Player 1 discussed above still cannot be \((0.05)\)-purified.

8 The Literature

Both approximate and exact purification appear in the literature. The underlying ideas are usually quite different; approximate theorems only require some form of continuity (the weakest being the atomlessness assumed here), whereas exact theorems usually assume, in addition, a combination of independence and finiteness.

Bellman and Blackwell [1949] and Dvoretzky, Wald, and Wolfowitz [1951] pioneered the area. Both treated two-person games in which only Player 1's observations and actions are explicit, Player 2 being represented directly by his strategies (compare Aumann [1964], which also treats games that are "extensive" for one player, "normal" for all others). Bellman and Blackwell use the alternate-interval idea described at the beginning of §3 to purify 1's strategies approximately when the payoff satisfies certain continuity conditions.\(^{13}\) Dvoretzky, Wald, and Wolfowitz use Lyapunov's theorem on the range of a vector measure to purify 1's strategies exactly when 2 has only finitely many strategies; in the opposite case, they adduce an example showing that exact purification is in general impossible. Dvoretzky, Wald, and Wolfowitz also prove an approximate purification theorem when 2's strategy space is conditionally compact\(^{14}\) in a metric based on the payoffs; the proof uses the conditional compactness to approximate 2's strategy space by a finite space, and then applies the exact purification theorem quoted above.

In the more recent literature, as in this paper, all players' observations and actions are explicit. Radner and Rosenthal [1982] purify equilibria exactly when the observations have independent, atomless distributions, and each player's payoff depends only on his own observation\(^ {15}\); whereas Milgrom and Weber show that equilibria can still be exactly purified when there is a finite family of mutually exclusive and exhaustive events, conditional on each of which the condition of Radner and Rosenthal

\(^{13}\) Piecewise continuity in 1's observations (using the metric of the unit interval), uniform over 1's actions and 2's strategies.

\(^{14}\) Every sequence has a Cauchy subsequence.

\(^{15}\) Radner and Rosenthal also obtain a somewhat more general theorem on exact purification, which is still not as general as that of Milgrom and Weber.
holds. Milgrom and Weber also purify strategies approximately when the joint distribution \( \mu \) of observations is absolutely continuous with respect to the product of the marginals (i.e., the distributions of individual observations), each of which they assume atomless. Here again the alternate-interval idea works; thus by transforming the marginals to Lebesgue measure, one can assume that \( \mu \) is absolutely continuous w.r.t. Lebesgue measure, and then the \( S'' \) form a purifying sequence (cf. §3).

While determining the precise relationships between all these results is not always a straightforward matter, it may be seen that our results are not subsumed under any of them. In the other direction, it should be noted that Milgrom and Weber work with compact action spaces, rather than the finite action spaces we assumed; but for finite action spaces, our Corollary A implies the approximate purification theorem of Milgrom and Weber. Finally, both the papers of Milgrom and Weber and of Radner and Rosenthal discuss topics unrelated to purification, including conditions for the existence of (approximate or exact) equilibria.

References


