

Switching Costs in Infinitely Repeated Games¹

Barton L. Lipman²
Boston University

Ruqu Wang³
Queen's University

Current Draft

September 2001

¹The authors thank Ray Deneckere for making us aware of related work.

²Email: blipman@bu.edu.

³Email: wangr@qed.econ.queensu.ca

Abstract

We show that small switching costs can have surprisingly dramatic effects in infinitely repeated games if these costs are large relative to payoffs in a single period. This shows that the results in Lipman and Wang [2000] do have analogs in the case of infinitely repeated games.

1 Introduction

In recent work, Lipman and Wang [2000], we showed that switching costs can have surprisingly strong effects in frequently but finitely repeated games. More specifically, suppose we have a finite stage game with actions chosen every Δ instants with a total length of time of play equal to M . Suppose the payoff to player i from the sequence of action profiles (a^1, \dots, a^T) is given by

$$\sum_{t=1}^T [\Delta u_i(a^t) - \varepsilon I_i(a^t, a^{t-1})]$$

where $I_i(a, a') = 0$ if $a_i = a'_i$ and 1 otherwise. In other words, he receives the payoff associated with a given action vector times the length of time these actions are played, minus a cost for each time he himself changes actions. We showed some very unexpected behavior in such games for small ε and Δ as long as ε is sufficiently large relative to Δ . For example, in games like the Prisoners' Dilemma which have a unique subgame perfect equilibrium outcome without switching costs, we obtain multiple equilibrium outcomes. In other games, such as coordination games, which have multiple equilibria without switching costs, we showed that one can have a unique subgame perfect equilibrium with small switching costs.

The analysis used finite repetition in a critical way. We noted that if the switching cost is large relative to one period worth of payoff, then no player would find it worthwhile to change actions in the last period regardless of what actions were played in the preceding period. This causes the usual backward induction arguments to break down. The fact that actions must be fixed at the end can have large effects early in the game. It is hard to see how a similar effect could be obtained in an infinitely repeated game.

Here we show that different but also surprising results are possible with switching costs in infinitely repeated games. Again, the key is whether the switching cost is large relative to one period's worth of payoff. To understand this, suppose we consider a simple variation on the usual discounting formulation, evaluating paths of play by the discounted sum over periods of the payoff in a period minus a switching cost if incurred in that period. More specifically, suppose player i 's payoff to a sequence of action profiles a^1, a^2, \dots is

$$(1 - \delta) \sum_{t=1}^{\infty} \delta^{t-1} [u_i(a^t) - \varepsilon I_i(a^t, a^{t-1})] \quad (1)$$

where $I_i(a, a') = 0$ if $a_i = a'_i$ and 1 otherwise as before. This formulation gives no obvious way to shrink the switching cost while maintaining the relationship between the payoff in a period and the switching cost. The only way in which period length can be thought of as entering this formulation is through the discount rate δ , which affects game payoff

and switching costs in the same way. Hence if we reduce ε , we *must* reduce it relative to $u_i(a)$ and thus relative to one period worth of payoff.

To see why this formulation is not obviously appropriate, suppose we think in terms of the formulation sketched above for the finite case where actions can be changed only at intervals of Δ instants. Think of the stage game payoffs as flow rates. It seems only reasonable to view the switching cost as an immediate payment, not a flow cost. Hence it seems natural to say that the way an agent would evaluate the payoff to the infinite sequence of actions above would be

$$\sum_{t=0}^{\infty} \int_{t\Delta}^{(t+1)\Delta} e^{-rs} u_i(a^t) ds - \sum_{t=0}^{\infty} e^{-rt\Delta} \varepsilon I_i(a^{t-1}, a^t).$$

If we integrate this, normalize $r = 1$, and set $\delta = e^{-\Delta}$, we get

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(a^t) - \sum_{t=0}^{\infty} \delta^t \varepsilon I_i(a^{t-1}, a^t).$$

Now we can reduce the switching cost and simultaneously maintain the relationship between the switching cost and one period worth of payoff: we increase δ toward 1 as we reduce ε . Note that increasing δ toward 1 corresponds exactly to reducing Δ , the length of a period, toward 0, just as we did in the finite horizon case. Let $G(\varepsilon, \delta)$ denote the infinitely repeated game where players evaluate payoffs using this criteria. We will think of $G(\varepsilon, 1)$ as the case where players use the limit of this function as $\delta \rightarrow 1$ — that is,

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} u_i(a^t) - \varepsilon \#\{t \mid a_i^t \neq a_i^{t-1}\}$$

where $\#$ denotes cardinality. We characterize equilibrium payoffs of $G(\varepsilon, \delta)$ for ε near or equal to 0 and δ near or equal to 1. Let $\mathcal{U}(\varepsilon, \delta)$ denote the set of equilibrium payoffs of $G(\varepsilon, \delta)$.

In line with the intuition suggested above, our results show that the set of equilibrium payoffs is exactly the usual Folk Theorem set if the switching cost is small relative to a period worth of payoff. In other words,

$$\lim_{\delta \uparrow 1} \left[\lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, \delta) \right] = \lim_{\delta \uparrow 1} \mathcal{U}(0, \delta),$$

the usual Folk Theorem payoffs. On the other hand, if switching costs are large relative to one period worth of payoff, this is not true. We have two results illustrating this point.

First, we consider the result of taking the two limits in the reverse order. In this case, there are two differences between the limiting payoff set and the usual Folk Theorem set. First, the relevant minmax payoff is now the pure minmax, not the usual minmax.

Intuitively, if a player needs to randomize to avoid punishment, the expected costs of switching actions makes this too costly. Second, the notion of feasibility changes as well since the switching costs can dissipate payoffs even in the limit as $\varepsilon \downarrow 0$. For example, in the coordination game¹

	a	b
a	3, 3	0, 0
b	0, 0	1, 1

the usual Folk Theorem set is all payoff vectors (u_1, u_2) where $u_1 = u_2$ and $.75 \leq u_i \leq 3$. By contrast, as $\delta \rightarrow 1$ and then $\varepsilon \downarrow 0$, the set of equilibrium payoffs converges to the set of all (u_1, u_2) such that $(0, 0) \leq (u_1, u_2) \leq (3, 3)$.

Second, we consider $\lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, 1)$ — that is, consider the set at the limit in δ and then take the limit as $\varepsilon \downarrow 0$. In this case, the switching cost is larger than the payoffs for *any* finite number of periods, so, naturally, we would expect the cost to have the largest effect here. In fact, both of the two earlier differences between the equilibrium payoff set and the usual Folk Theorem set remain and a third is added. This third difference is strikingly unusual: payoffs that are supported by putting some weight on payoff vectors that are not individually rational cannot be obtained. For example, in the Prisoners’ Dilemma,

	C	D
C	3, 3	0, 5
D	5, 0	1, 1

the usual Folk Theorem set is all feasible payoffs where each player gets at least 1. When we set $\delta = 1$ and take the limit as $\varepsilon \downarrow 0$, the set of equilibrium payoffs converges to the set of payoffs where each player gets at least 1 and neither gets more than 3. We get this result because in this game, we cannot put any “weight” on the $(5, 0)$ or $(0, 5)$ payoff vector. (Payoff vectors which are not convex combinations of $(3, 3)$ and $(1, 1)$ are obtained by players dissipating payoffs through the switching costs.) Intuitively, with this formulation, any path of play in the game must have the property that players change actions only finitely often. Hence any path eventually “absorbs” in the sense that at some point, actions never change again. It is obvious that we cannot have an equilibrium where the players know that actions will never change again from (C, D) or (D, C) since the player getting 0 will change actions. What is less obvious is why we cannot have some kind of randomization that “hides” from the players the fact that no further changes of action will occur. (We do allow the players to condition on public randomizing devices, so we give the maximum possible ability for the players to use such strategies.) We show that players must eventually become sure enough that no change will occur that they will deviate from any such proposed equilibrium.

Our results differ from those of Chakrabarti [1990] who considers a similar model.

¹This game violates the genericity assumptions we will use in our analysis. However, it is not hard to show that our results carry over to this game.

He analyzes infinitely repeated games with a more general “inertia cost” than we consider. His payoff criterion, however, does not fit into the class we consider. Specializing his switching cost to our setting, he assumes players evaluate payoffs to the sequence (a^1, a^2, \dots) by

$$\liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} [u_i(a^t) - \varepsilon I_i(a^{t-1}, a^t)],$$

the $\delta = 1$ version of (1) above. Thus he cannot let ε go to zero while maintaining the relationship between switching cost and payoff in a period. His formulation is a special case of a dynamic game as considered by Dutta [1995], while our formulation is not. Using the results of Dutta [1995], one can show that Chakrabarti’s set of equilibrium payoffs differs from the usual Folk Theorem set in two ways: individual rationality must be defined using the pure minmax and feasibility must be defined to allow the possibility that switching costs dissipate some payoffs.² However, the third effect we obtain is not present.

In the next section, we state the model. In Section 3, we give our characterizations of equilibrium payoffs. In the final section, we offer some concluding remarks. Proofs are contained in the Appendix.

2 Model

Fix a finite stage game $G = (A, u)$ where $A = A_1 \times \dots \times A_I$, each A_i is finite and contains at least two elements, and where $u : A \rightarrow \mathbf{R}^I$. Let S_i denote the set of mixed stage game strategies — that is, S_i is the set of randomizations over A_i . We allow the players to use public randomizing devices, so a strategy can depend on the history of play as well as the outcome of the public randomization. For simplicity, we will suppose that there is an iid sequence of random variables, ξ_t , which are uniformly distributed on $[0, 1]$ which all players observe. A strategy for player i , then, is a function from the history of past actions and the realization of the randomizations (up to and including the current period) into S_i . The payoffs in the game $G(\varepsilon, \delta)$ are defined as follows. Given a sequence of actions (a^1, a^2, \dots) where $a^t = (a_1^t, \dots, a_I^t)$, i ’s payoff given this sequence is

$$(1 - \delta) \sum_{t=0}^{\infty} \delta^t u_i(a^t) - \sum_{t=0}^{\infty} \delta^t \varepsilon I_i(a^{t-1}, a^t).$$

where $I_i(a^{t-1}, a^t) = 1$ if $a_i^{t-1} \neq a_i^t$ and 0 otherwise. The payoffs in the game $G(\varepsilon, 1)$ from this sequence are

$$\left[\liminf_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T u_i(a^t) \right] - \varepsilon \#\{t \mid a_i^t \neq a_i^{t-1}\}$$

²This is not how Charkabarti states his result.

where $\#$ denotes cardinality. Let $\mathcal{U}(\varepsilon, \delta)$ denote the closure of the set of subgame perfect equilibrium payoffs in $G(\varepsilon, \delta)$.

Remark 1 As shown by Fudenberg and Maskin [1991], the use of public randomization is purely a matter of convenience in the usual repeated game. More specifically, one can replace the public randomization by constructing an appropriate deterministic sequence of actions. However, the assumption is not as innocuous here. While it is not needed for any of the other results, the result of Theorem 4 is not true in general without public randomization. The reason is simply that the deterministic sequence of actions that would replace the public randomization in the usual argument will typically require numerous changes of action. Since this is costly here, such behavior can be difficult to support as an equilibrium. On the other hand, the most interesting aspect of Theorem 4 is the payoffs which *cannot* be achieved. Since allowing public randomization can only increase the set of equilibrium payoffs, this is the most interesting case to consider for that result.

We define limits of $\mathcal{U}(\varepsilon, \delta)$ in the following way. Say that $u \in \lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, \delta)$ if there is a sequence $\varepsilon_n \rightarrow 0$ with $\varepsilon_n > 0$ for all n and a sequence $u^n \rightarrow u$ such that $u^n \in \mathcal{U}(\varepsilon_n, \delta)$ for all n . Define $\lim_{\delta \uparrow 1} \mathcal{U}(\varepsilon, \delta)$ in the analogous fashion. Similarly,

$$u \in \lim_{\varepsilon \downarrow 0} \left[\lim_{\delta \rightarrow 1} \mathcal{U}(\varepsilon, \delta) \right]$$

if there are sequences $\varepsilon_n \rightarrow 0$ and $u^n \rightarrow u$ such that $\varepsilon_n > 0$ for all n and

$$u^n \in \lim_{\delta \rightarrow 1} \mathcal{U}(\varepsilon_n, \delta), \quad \forall n.$$

The limit in the reverse order is defined symmetrically.

Define i 's *minmax payoff*, v_i , by

$$v_i = \max_{s_i \in S_i} \min_{s_{\sim i} \in S_{\sim i}} u_i(s_i, s_{\sim i}).$$

Let

$$R = \{u \in \mathbf{R}^I \mid u \geq v\}$$

denote the usual set of individually rational payoffs where $v = (v_1, \dots, v_I)$. Define i 's *pure minmax payoff*, w_i , by

$$w_i = \max_{a_i \in A_i} \min_{s_{\sim i} \in S_{\sim i}} u_i(a_i, s_{\sim i}).$$

Let

$$W = \{u \in \mathbf{R}^I \mid u \geq w\}$$

denote what we will call the set of *weakly individually rational* payoffs, where $w = (w_1, \dots, w_I)$.³ Note that $w_i \leq v_i$ for all i so $R \subseteq W$. Note also that without loss of generality, we can assume that the actions by the players other than i which minimize i 's payoff in the pure minmax case are pure. We exploit this fact in what follows.

For any set $B \subseteq \mathbf{R}^I$, let $\text{conv}(B)$ denote its convex hull. Let U denote the set of payoffs feasible from pure strategies and let F denote the usual set of feasible payoffs. That is,

$$U = \{u \in \mathbf{R}^I \mid u = u(a), \text{ for some } a \in A\}$$

and $F = \text{conv}(U)$. For comparison purposes, we first state the usual Folk Theorem.

Theorem 1 (The Folk Theorem)

A. Assume that for all i , there is $u^i \in F \cap R$ such that $u^i > v_i$. Then

$$\mathcal{U}(0, 1) = F \cap R.$$

B. Assume that F has dimension I . Then

$$\lim_{\delta \uparrow 1} \mathcal{U}(0, \delta) = F \cap R.$$

This result is a trivial extension of theorems in Fudenberg and Tirole [1991], Chapter 5, and so we omit the proof.

Remark 2 The assumption stated in Theorem 1.A is not standard. We use it to ensure that the set of feasible payoffs with every player i receiving strictly more than v_i is nonempty. Since the usual version of the undiscounted Folk Theorem considers such payoffs, this enables us to use that result most easily. For part B, we use the most simply stated sufficient condition. It could be replaced by the weaker NEU condition of Abreu, Dutta, and Smith [1994].

3 Results

For simplicity, our results all assume that G is *generic* in the sense that the following two conditions hold. First, for each player i , there is a unique vector of actions a^i such that $w_i = u_i(a^i)$. Second, $u_i(a_i, a_{\sim i}) = u_i(a'_i, a_{\sim i})$ if and only if $a_i = a'_i$.

³Given vectors x and y , we use $x \geq y$ to mean greater than or equal to in every component and $x \gg y$ to mean strictly larger in every component.

The easiest set to characterize is $\lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, \delta)$. Once ε becomes small enough that even an infinite number of changes of action is worthwhile for a tiny one period payoff gain, clearly, the switching cost ceases to be relevant. Formally, we have

Theorem 2 *For any generic game such that F has dimension I ,*

$$\lim_{\delta \uparrow 1} \left[\lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, \delta) \right] = F \cap R.$$

That is, if we take the limit as $\varepsilon \downarrow 0$ first, we obtain the set of feasible, individually rational payoffs.

The proof of this result is a simple extension of the proof in Fudenberg and Tirole and so is omitted. The argument is to note that their proof for the observable mixed strategy case involves *strict* payoff comparisons. Hence small enough switching costs cannot affect the optimality of the strategies in question. Since the public randomization effectively creates observable mixed strategies, this completes the argument. We conjecture that one could extend the usual arguments in the case where mixed strategies are not observable by adapting the randomizations constructed to take account of the switching cost.

As discussed in the introduction, there are two ways one can consider the case where switching costs are large relative to one period of payoff. The first is to take limits as $\delta \uparrow 1$ first and then as $\varepsilon \downarrow 0$. The second is to set $\delta = 1$ and consider the limit as $\varepsilon \downarrow 0$. The former is the simpler to characterize.

Given a set $B \subseteq \mathbf{R}^I$, let $c(B)$ denote the comprehensive, convex hull of B . That is, $c(B)$ is the set of points less than or equal to a convex combination of points in B . Define $F^* = c(U)$. This is the feasible set of payoffs when we allow players the ability to “throw away” utility.

Theorem 3 *For any generic game,*

$$\lim_{\varepsilon \downarrow 0} \left[\lim_{\delta \uparrow 1} \mathcal{U}(\varepsilon, \delta) \right] = F^* \cap W.$$

That is, the limiting payoff set is the set of feasible payoffs (taking into account the ability to dissipate payoffs by switching actions) which are weakly individual rational.

The proof is in the Appendix, but here we sketch the idea. Suppose we wish to construct strategies generating a payoff in the set above. A natural way to do so is to begin by having each player switch actions some number of times to dissipate the

appropriate amount of utility and then have the players follow a cycle which generates a certain payoff in $\text{conv}(U)$. Ignoring the discreteness in the switching cost, any payoff in F^* can be generated this way. Suppose we lengthen the cycle by having the players use each action vector for a larger number of consecutive periods. Clearly, this makes changes of action during the cycle rarer and rarer. By making the cycle longer as we increase δ , we are able to keep these switching costs relatively small even as $\delta \uparrow 1$. On the other hand, consider the payoff a player receives if the others are trying to minimize his payoff. If they continually move to the action which minimizes his payoff given the action he has most recently played, he will either stop changing actions and get his pure minmax or change actions every period. As $\delta \uparrow 1$, these switching costs are exploding, ensuring that this option is suboptimal. Hence, if need be, we can force a player down to his pure minmax payoff. Given these two facts, the rest of the proof is similar to a standard Folk Theorem construction.

The most complex case is when we set $\delta = 1$ and consider the limit as $\varepsilon \downarrow 0$. Let U_{\geq} denote those points in U which are greater than w . That is,

$$U_{\geq} = \{u \in U \mid u \geq w\}.$$

For this result, we need one additional assumption which we call *rewardability*: we assume that for every player i , there is a $u \in U_{\geq}$ with $u_i > w_i$. To see the idea behind the name, suppose this does not hold. As mentioned in the introduction, when $\delta = 1$, the only u vectors which can be achieved infinitely often with positive probability are those in U_{\geq} . If for some player i , all these vectors give him w_i , then he cannot be rewarded for aiding in the punishment of a deviator. This complication restricts the set of equilibria in a complex fashion. Similarly to Wen [1994], one can explicitly work out the way in which this constrains punishments to give an exact characterization of the limiting equilibrium set without this restriction.

Theorem 4 *For any generic game satisfying rewardability,*

$$\lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, 1) = c(U_{\geq}) \cap W.$$

To see how this differs from the payoff set from Theorem 3, note that we can write that set as

$$\lim_{\varepsilon \downarrow 0} \left[\lim_{\delta \uparrow 1} \mathcal{U}(\varepsilon, \delta) \right] = c(U) \cap W.$$

In this form, the difference is obvious: the payoff set of Theorem 4 only puts weight on payoffs which are weakly individually rational, not all pure strategies.

The proof that any payoff in $c(U_{\geq}) \cap W$ is a limiting equilibrium payoff is similar to standard Folk Theorem arguments. The more unusual part of the proof is the demonstration that no payoff outside this set can be close to an equilibrium payoff. We sketch

the idea in the context of the Prisoners' Dilemma we used in the introduction:

	C	D
C	3, 3	0, 5
D	5, 0	1, 1

Let $\mathcal{P}(C, C)$ denote the set of infinite sequences of actions which eventually “absorb” at (C, C) — that is, for which at some point in time, play reaches (C, C) and never changes again. Define $\mathcal{P}(C, D)$, etc., analogously. Note that any sequence of actions which is not in $\mathcal{P}(C, C)$, $\mathcal{P}(C, D)$, $\mathcal{P}(D, C)$, or $\mathcal{P}(D, D)$ has the players changing actions infinitely often. If any player has a positive probability of switching actions infinitely often, his expected payoff is $-\infty$ and so his strategy cannot be optimal. Hence any equilibrium has to put zero probability on such an event. That is, the other four sets must have probability 1 in total. The main claim of Theorem 4 is that the sets $\mathcal{P}(C, D)$ and $\mathcal{P}(D, C)$ must also have zero probability in equilibrium.

To see this, suppose, say, $\mathcal{P}(C, D)$ has probability $\mu > 0$. Suppose player 1 updates his beliefs over the path of play conditional on the event that (C, D) is the action profile played at period t . Clearly, any path of play which absorbs at a different action profile at a period before t must have zero probability at this point. As t gets large, the set of paths being eliminated this way gets closer to the set of all paths not absorbing at (C, D) . At the same time, this event cannot rule out the possibility that play has already absorbed at (C, D) . In fact, as t gets large, the set of paths where play absorbs at (C, D) at some date prior to t is converging to $\mathcal{P}(C, D)$. Hence as t gets large, the conditional probability on $\mathcal{P}(C, D)$ is going to 1. But once this conditional probability is large enough, player 1 will certainly deviate to D , a contradiction. Note that this argument actually implies that we cannot have a *Nash* equilibrium putting positive probability on $\mathcal{P}(C, D)$, much less a subgame perfect equilibrium.

It is natural to wonder why we get such a dramatic difference between the $\delta = 1$ and $\delta \uparrow 1$ cases. This is much more than the dimensionality issue that comes up in the case of repeated games without switching costs. The difference here hinges, as with most of our results, on the relationship between the switching cost and the length of a period. As noted, when we take the limit as $\delta \uparrow 1$ before taking the limit as $\varepsilon \downarrow 0$, there is a natural sense in which we are making the period length short relative to the switching cost. However, this effect can be undone in equilibrium. To see the point, note that we could always construct equilibria in which the players act as if a block of k periods was only one period. That is, they only change actions at intervals of k periods. Of course, one cannot prevent players from deviating from this and changing actions more frequently. However, the punishment for deviations can also come more quickly as well. By constructing such equilibria, we can effectively make the length of a period arbitrarily long relative to the switching cost.

For any $\delta < 1$, this matters. However, for $\delta = 1$, it does not. When $\delta = 1$, only

the number of times the players change actions matters, not the intervals at which these changes occur. Hence this is the only situation where the players cannot endogenously alter the relationship between switching costs and payoffs in “a period.”

4 Conclusion

In light of these results, we see that the limiting set of equilibrium payoffs depends on the order of limits and whether we fix one parameter at the limit. It is worth emphasizing that three sets of payoffs discussed in our three theorems can be very different, so that the discontinuity identified is nontrivial. See the examples in the introduction for illustration. As stressed throughout, we see these results as indicating that even small switching costs can play a significant role in determining equilibrium outcomes as long as these costs are large enough relative to a period worth of payoff.

A Proof of Theorem 3

It is obvious that no $u \notin F^* \cap W$ can be an equilibrium payoff. Such a u is either infeasible or has some player with a lower payoff than what he could guarantee himself by a constant action. Hence we only need to show that every $u \in F^* \cap W$ is in the limiting set of equilibrium payoffs.

First, suppose that for some player i , every $u \in F^* \cap W$ has $u_i = w_i$. By our genericity assumption, there is a unique action profile a^i such that $u_i(a^i) = w_i$. Hence the only way this can happen is if $F^* \cap W = \{u(a^i)\}$. On the other hand, consider the payoff vector when all players play their minmax action. By definition, if i plays his minmax action, his payoff must be at least w_i . Hence the resulting payoff vector, say \hat{u} , must satisfy $\hat{u} \geq w$. Obviously, \hat{u} is feasible, so $\hat{u} \in F^* \cap W$. Hence it must be true that $\hat{u} = u(a^i)$. That is, a^i is the vector of minmax actions. In this case, a^i must be a Nash equilibrium of the stage game. To see this, suppose any player j deviates. Consider any other player k . Since k is using his minmax action, his payoff must be at least w_k . Suppose j strictly gains from the deviation. Then the payoff vector at the new action profile is larger than but not equal to w and hence is also in $F^* \cap W$, a contradiction. Given this, it is not hard to see that a subgame perfect equilibrium of the repeated game is for all players to play their minmax actions in every period. This generates payoff w , just as the theorem states.

Now suppose that for every player i , there is a $u^i \in F^* \cap W$ which has $u^i_i > w_i$. Since $u^i_j \geq w_j$ by $u^i \in W$, we have $\sum_i \alpha_i u^i \gg w$ for any strictly positive α_i 's summing to 1. Hence there must be a $u \in F^* \cap W$ with $u \gg w$.

In light of this, pick any $u' \in F^* \cap W$ with $u' \gg w$. We will show that any such $u' \in \lim_{\varepsilon \downarrow 0} [\lim_{\delta \uparrow 1} \mathcal{U}(\varepsilon, \delta)]$. Because any $u'' \in F^* \cap W$ with $u'' \not\gg w$ can always be obtained as a limit of a sequence of such u' , this will complete the proof.

By definition of F^* , there exist strictly positive numbers $\alpha_1, \dots, \alpha_Z$ and pure actions a^1, \dots, a^Z , such that $\sum_z \alpha_z = 1$, and

$$u' \leq \sum_z \alpha_z u(a^z).$$

Define m_i to be the largest integer m such that

$$m\varepsilon \leq \sum_z \alpha_z u_i(a^z) - u'_i$$

and

$$m\varepsilon \leq \sum_z \alpha_z u_i(a^z) - w_i - 6\varepsilon - 2Z\varepsilon.$$

(Choose ε sufficiently small that this is well-defined.) Note that this m_i depends on ε , but not on δ . Because $u'_i > w_i$ by assumption, for ε sufficiently small, the binding constraint will be the first inequality. Hence as $\varepsilon \downarrow 0$, $m_i \varepsilon \rightarrow \sum_z \alpha_z u_i(a^z) - u'_i$.

Fix any $\eta \in (0, 1)$. Define k_1 through k_Z by

$$\eta^{k_z} = \frac{2 - \sum_{j=1}^z \alpha_j}{2 - \sum_{j=1}^{z-1} \alpha_j}, \quad z = 1, \dots, Z$$

where we define $\sum_{j=1}^0 \alpha_j = 0$. It is easy to see that the right-hand side is strictly between 0 and 1 for all z so k_z is well defined. It is not hard to show that this definition implies

$$\eta^{\sum_{j=0}^{z-1} k_j} (1 - \eta^{k_z}) = \frac{\alpha_z}{2}.$$

Note also that

$$\sum_{z=1}^Z \eta^{\sum_{j=0}^{z-1} k_j} (1 - \eta^{k_z}) = 1 - \eta^K$$

where $K = \sum_z k_z$. Also, $\sum_z \alpha_z / 2 = 1/2$. Hence $1 - \eta^K = 1/2$.

Without loss of generality, we can assume that the k_z 's are all integers. To see this, first suppose there is a solution where all k_z 's are rational numbers. Clearly, we can find a common denominator and write $k_z = \gamma_z / \Gamma$ for all z where the γ_z 's and Γ are integers. But then let $\bar{\eta} = \eta^{1/\Gamma}$ and use $\bar{\eta}$ in place of η . The associated k_z 's will then be the γ_z 's and hence will be integers. So suppose there is no solution where all the k_z 's are rational. Then we can approximate the solution arbitrarily closely with a rational solution. Since we will be constructing a sequence of payoffs, we simply improve our approximation through the sequence and obtain convergence.

Let $\delta_n = \eta^{1/n}$. Clearly, as $n \rightarrow \infty$, $\delta_n \rightarrow 1$. Consider the strategies where players first switch actions back and forth, with i stopping after he has changed actions m_i times and ending up at a_i^1 . After all players have completed this, they play a^1 nk_1 times, then a^2 nk_2 times, etc., repeating the cycle forever.

Player i 's payoff along this path is

$$(1 - \delta_n) \sum_{t=0}^M \delta_n^t u_i(\bar{a}^t) - \sum_{t=0}^{m_i-1} \delta_n^t \varepsilon + \delta_n^M \left\{ \sum_{z=1}^Z \left[u_i(a^z) (1 - \delta_n^{nk_z}) - \varepsilon I_i(z) \right] \frac{\delta_n^{\sum_{j=0}^{z-1} nk_j}}{1 - \delta_n^K} \right\} + \delta_n^M \varepsilon I_i(1)$$

where $M = \max_i m_i$, \bar{a}^t is the action played in period t during the switching phase, and $I_i(z)$ is 1 if $a_i^z \neq a_i^{z-1}$ and 0 otherwise. In this expression, we use the conventions that $a^{-1} = a^Z$ and $\sum_{j=0}^0 nk_j = 0$.

Substituting for δ_n in terms of η gives

$$(1 - \eta^{1/n}) \sum_{t=0}^M \eta^{t/n} u_i(\bar{a}^t) - \sum_{t=0}^{m_i-1} \eta^{t/n} \varepsilon + \eta^{M/n} I_i(1) \varepsilon + \eta^{M/n} \left\{ \sum_{z=1}^Z \left[u_i(a^z)(1 - \eta^{k_z}) - \varepsilon I_i(z) \right] \frac{\eta^{\sum_{j=0}^{z-1} k_j}}{1 - \eta^{k_z}} \right\}$$

Let $u_i^*(n, \varepsilon)$ denote this expression. Note that the term in brackets is independent of n . Substituting using our definition of the k_z 's, this is

$$(1 - \eta^{1/n}) \sum_{t=0}^M \eta^{t/n} u_i(\bar{a}^t) - \sum_{t=0}^{m_i-1} \eta^{t/n} \varepsilon + \eta^{M/n} I_i(1) \varepsilon + \eta^{M/n} \left\{ \sum_{z=1}^Z \left[u_i(a^z) \alpha_z - \varepsilon I_i(z) \frac{\alpha_z}{1 - \eta^{k_z}} \right] \right\}$$

As $n \rightarrow \infty$, this converges to

$$-m_i \varepsilon + \varepsilon I_i(1) + \sum_{z=1}^Z \left[u_i(a^z) \alpha_z - \varepsilon I_i(z) \frac{\alpha_z}{1 - \eta^{k_z}} \right].$$

As $\varepsilon \downarrow 0$, this converges to u_i' .

Also, note that the definition of m_i implies that the expression above is at least

$$w_i - \sum_{z=1}^Z \alpha_z u_i(a^z) + 6\varepsilon + 2Z\varepsilon + \sum_{z=1}^Z \alpha_z u_i(a^z) - \sum_{z=1}^Z \varepsilon I_i(z) \frac{\alpha_z}{1 - \eta^{k_z}}.$$

From the definition of k_z , we must have $\alpha_z/(1 - \eta^{k_z}) \leq 2$ for all z . Hence this is at least $w_i + 6\varepsilon$. In light of this, for any ε sufficiently small, there is a \bar{n} such that $u_i^*(n, \varepsilon) > w_i + 5\varepsilon$ for all $n \geq \bar{n}$.

To complete the specification of the strategies, we need to specify what happens in response to a deviation. To explain this most simply, we divide the histories of the game into what we call *normal mode* and two kinds of *punishment modes*. For each punishment mode, there is a unique player i whom we refer to as the *target* of the punishment. These are defined as follows. First, if the history of the game is such that there have been no deviations from the path described above, then we are in normal mode. Any history where the first deviation occurred in the previous period is in a type 1 punishment mode and the target is that player who deviated. (In case of multiple simultaneous deviations, choose any subgame equilibrium. Naturally, such histories will not affect equilibrium considerations.) When in any punishment mode, the target is supposed to change actions four times, ending up at the action he took in the period he deviated. The following period, he is supposed to play the action he should have played in the period he deviated and then follows the equilibrium path from there. While he does this, the other players do not change actions. If the target does not carry this out (either by not changing actions when he is supposed to or by ending up at the wrong action after his four changes), all players other than the target choose that vector of actions which

minimize the target's payoff on the assumption that the target continues with the action he played in the previous period. We then continue as above. That is, we wait for the target to change actions four times (not counting whatever changes of action preceded the deviation from the punishment) and end up at the most recent deviation action. After this, all players use the actions that were to have been played in the period in the period in which the original deviation took place. At this point, we are back to normal mode and follow the equilibrium path from there. Again, if the target deviates, the other players change actions to the appropriate minimizing actions and we begin again.

If any player other than the target deviates while in punishment mode, we move to the second type of punishment mode with the deviator as the target. As with the first type of punishment mode, the punishment works by having the target change actions four times returning to the deviation action enforced by having all players other than the target minimize the target's payoff under the hypothesis that the target will play the same action as the previous period if the target does not carry this out. The only difference between the two types of punishment mode concerns what happens when the target returns to the action he played when he deviated. Instead of returning to where we were prior to this deviation (i.e., returning to the punishment of the previous target), we treat this as a return to the equilibrium path. So in the subsequent period, all players return to the actions they would have played in the period after the original deviation had no deviation occurred. We again return to normal mode. As before, the target's strategy in such a punishment mode is to switch four times, ending at the deviation action in order to return play to normal mode.

To see that the strategies described above form a subgame perfect equilibrium for any sufficiently large n , consider any history. First, suppose we are in either type of punishment mode and i is the target. Let \bar{u}_i be i 's highest possible payoff in the stage game. The options that could be optimal for i are: never change actions again, change to the minmax action and stay there forever, change actions every period to "manipulate" the changing of the others, or to carry out his four changes of action and return to the equilibrium path. The first yields $(1 - \delta_n)\bar{u}_i + \delta_n w_i$ at best. The second yields no more than $(1 - \delta_n)\bar{u}_i + \delta_n w_i - \varepsilon$. Obviously, for n sufficiently large, the former is very close to w_i and the latter to $w_i - \varepsilon$. The third option gives i , at most, $\bar{u}_i - \varepsilon / (1 - \delta_n)$. As n gets large, this payoff goes to $-\infty$. Clearly, then, for n sufficiently large, if i does not return to the equilibrium path, his payoff is approximately w_i or less. If i returns to the equilibrium, then for n sufficiently large, an approximate lower bound for his payoff is $u_i^*(n, \varepsilon) - 5\varepsilon$. To see this, recall that he must change actions four times during the punishment and then ends up at the profile where he deviated. At this point, he must change actions a fifth time to bring play back to the equilibrium. To see why this is an approximate lower bound rather than an approximation of the payoff itself, note that if i had carried out some of his m_i changes of actions before deviating, then the continuation payoff after returning to the equilibrium path is strictly larger than $u_i^*(n, \varepsilon)$. Also, if i has already

deviated from the punishment, when he goes back to the most recent deviation action, this might be the same as the action he is supposed to play in the next period, so he might not have to change actions the fifth time. By construction, $u_i^*(n, \varepsilon) - 5\varepsilon > w_i$ for n sufficiently large, so it is optimal for i to return to the equilibrium as specified above.

Second, suppose we are in normal mode. Consider player i . If he continues in normal mode, he expects some continuation payoff, say φ_i . If i deviates, we know that he will change actions four times and then return to the equilibrium path after that, requiring a fifth change of actions. Hence his payoff to deviating is approximately $\varphi_i - 5\varepsilon$. Clearly, if n is sufficiently large, i is better off continuing in normal mode.

Finally, suppose we are in either type of punishment mode and consider any player i who is *not* the target. We claim i cannot gain by deviating from the punishment. To see this, consider the worst case: where i has to change actions to punish the target because the target has deviated from the punishment. In this case, i expects to have to change actions at most twice, counting the return to the equilibrium path. To see this, recall that i is supposed to change actions to minmax the target. The target is expected to then carry out his required changes of action while i 's action remains constant. After this, we return to the equilibrium path, potentially requiring i to change actions again. Hence, letting φ_i denote i 's continuation payoff from the point where play returns to the equilibrium path, an approximate lower bound for i 's payoff to carrying out the punishment appropriately is $\varphi_i - 2\varepsilon$.

Suppose instead that i deviates. He will then be the target and, as shown above, will change actions at least four times. Hence an approximate upper bound for his payoff is $\varphi_i - 4\varepsilon$. So for n sufficiently large, i has no incentive to deviate. ■

B Proof of Theorem 4

First, fix any $\hat{u} \in c(U_{\geq}) \cap W$. By definition, $\hat{u} \geq w$. Also, there are action profiles, say a^1, \dots, a^Z , and strictly positive numbers $\alpha_1, \dots, \alpha_Z$ such that $u(a^z) \geq w$ for all z , $\sum_{z=1}^Z \alpha_z = 1$, and $\hat{u} \leq \sum_z \alpha_z u(a^z)$. Without loss of generality, assume $Z = 1$. Once we prove the result for this case, the fact that we allow public randomizations extends the result to larger Z .

Fix an $\varepsilon > 0$, smaller than any possible payoff difference. That is, choose ε so that

$$\varepsilon < \min_i \min_{u, u' \in u(A), u_i \neq u'_i} |u_i - u'_i|. \quad (2)$$

Genericity implies that there is no player whose payoff is constant over all $u \in u(A)$. Hence the right-hand side is strictly positive, so this is possible.

For each i such that $u_i(a^1) > w_i$, let c_i denote the largest nonnegative integer such that $u_i(a^1) - c_i\varepsilon \geq \hat{u}_i$. Since $u_i(a^1) \geq \hat{u}_i$, (2) implies that these must exist. Let u'_i denote $u_i(a^1) - c_i\varepsilon$ evaluated at this largest c_i . For i such that $u_i(a^1) = w_i$, we must have $\hat{u}_i = u_i(a^1)$ (since $\hat{u}_i \geq w_i$). For such i , define $c_i = 0$ and let $u'_i = w_i$. Clearly, as $\varepsilon \downarrow 0$, $u' \rightarrow \hat{u}$. In light of this, we show that $u' \in \mathcal{U}(\varepsilon, 1)$ for all sufficiently small ε , thus demonstrating $\hat{u} \in \lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, 1)$.

To show this, construct strategies as follows. In the first period, if c_i is even (where 0 is treated as even), player i plays a_i^1 . Otherwise, he plays any other action. For the next several periods, each player changes between a_i^1 and any other action, concluding when he has changed actions c_i times. At this point, by construction, he will be back to a_i^1 . Once all players have completed this phase, no player changes actions again, so a^1 is played forever after. It is easy to see that the payoffs if there are no deviations are u' .

To complete the specification of the strategies, we need to specify what happens in response to a deviation. Let \bar{u} be the equally weighted average of the payoff vectors in U_{\geq} — that is,

$$\bar{u} = \frac{1}{\#U_{\geq}} \sum_{u \in U_{\geq}} u.$$

Our rewardability assumption implies that $\bar{u} \gg w$.

For simplicity, we describe behavior at the out of equilibrium histories in terms of a number of different punishment modes. There is one punishment mode for each player and each action available to that player. So we refer to a typical punishment mode as the (i, a_i) punishment mode where $a_i \in A_i$. In punishment mode (i, a_i) , i is the target of the punishment and a_i is the action he played which started this punishment mode.

We go to punishment mode (i, a_i) if i is the first player to deviate from the equilibrium play above and deviates by playing a_i when he is supposed to play something different. (We ignore multiple simultaneous deviations throughout. Any specification of a subgame equilibrium will suffice for these histories.) If some player $j \neq i$ deviates while we are in punishment mode for i by playing action a_j when he is not supposed to, we move to punishment mode (j, a_j) . The reaction to deviations by player i during punishment mode (i, a_i) are explained as part of describing the mode.

In punishment mode (i, a_i) , all players other than i (the target) go to the actions which minimize the target's payoff under the hypothesis that the target plays the same action he played in the previous period. There is a number k_i of times that the target is supposed to change actions, independent of a_i . As long as the target continues to change actions, the other players do not change their actions. If the target stops changing before he has changed k_i times, the other players move to the actions which minimize the target's payoff under the hypothesis that he plays the action he stopped at. The exact sequence

of actions used by the target while changing actions is unimportant with two exceptions. First, the target's strategy is to change actions k_i times without stopping. Second, k_i will be even and the sequence must have the property that the target concludes the sequence by returning to a_i .

Once the target has changed actions k_i times, we have a publicly observed randomization to pick a vector from U_{\geq} . For any player i , let \underline{u}^i denote any $u \in U_{\geq}$ which minimizes u_i subject over U_{\geq} . The public randomization puts probability $q_i(a_i)$ on \underline{u}^i and with probability $1 - q_i(a_i)$ chooses uniformly from (all of) U_{\geq} . By genericity, each $u \in U_{\geq}$ is generated by a unique action profile. When the outcome of this randomization is observed, all agents change actions (if need be) to move to the associated action profile and never change actions again.

For the computation of k_i and $q_i(a_i)$, we need some more notation. Let $p_i(a_i)$ denote the probability that i will have to change actions again when the randomization is observed given that the randomization is uniform on U_{\geq} . That is, $p_i(a_i)$ is the probability that a_i is different from the action i plays in a uniformly drawn profile from U_{\geq} . (Recall that k_i is even and the target must end up at a_i at the end of his k_i changes of action.) Also, let $I_i(a_i) = 0$ if a_i is i 's minmax action and 1 otherwise. Similarly, let $\underline{I}_i(a_i)$ be 0 if a_i is the same action i plays at \underline{u}^i and 1 otherwise.

Let β_i be the smallest integer b such that $b\varepsilon \geq \underline{u}_i^i - w_i$. Note that $\underline{u}_i^i \geq w_i$, so this is well defined. Let k_i equal the smallest even integer greater than or equal to

$$\beta_i + 1 + 2 \max_{i,j|i \neq j} \frac{\bar{u}_i - \underline{u}_i^i + 1}{\bar{u}_j - w_j}.$$

Note that $\bar{u} \gg w$ implies that k_i is well defined. Set $q_i(a_i)$ so that

$$1 - q_i(a_i) = \frac{k_i\varepsilon + w_i - \underline{u}_i^i + [\underline{I}_i(a_i) - I_i(a_i)]\varepsilon}{\bar{u}_i - \underline{u}_i^i + [\underline{I}_i(a_i) - p_i(a_i)]\varepsilon}.$$

By construction, $k_i\varepsilon > \underline{u}_i^i - w_i + \varepsilon$, so the numerator is at least

$$\varepsilon + [\underline{I}_i(a_i) - I_i(a_i)]\varepsilon \geq 0.$$

For ε sufficiently small, the denominator must be strictly positive as well since $\bar{u}_i > \underline{u}_i^i$. Finally, as ε goes to zero, the fraction converges to 0, so it must be less than 1 for small enough ε . Hence $q_i(a_i)$ is well defined for ε sufficiently small.

The key fact to note about this choice of $q_i(a_i)$ is that it ensures that the target is indifferent between following the equilibrium punishment and not. To see this, note that the target's expected payoff to following the equilibrium punishment is

$$q_i(a_i)[\underline{u}_i^i - \underline{I}_i(a_i)\varepsilon] + [1 - q_i(a_i)][\bar{u}_i - p_i(a_i)\varepsilon] - k_i\varepsilon.$$

Rearranging, this is

$$\underline{u}_i^i - \underline{I}_i(a_i)\varepsilon - k_i\varepsilon + [1 - q_i(a_i)] \left\{ \bar{u}_i - \underline{u}_i^i + [\underline{I}_i(a_i) - p_i(a_i)]\varepsilon \right\}.$$

Substituting for $1 - q_i(a_i)$ from the above gives

$$\underline{u}_i^i - \underline{I}_i(a_i)\varepsilon - k_i\varepsilon + k_i\varepsilon + w_i - \underline{u}_i^i + [\underline{I}_i(a_i) - I_i(a_i)]\varepsilon = w_i - I_i(a_i)\varepsilon.$$

Suppose that i does not follow the equilibrium punishment. What is the best alternative? Clearly, i can either not change actions ever again or change to his minmax action (if he is not already playing it) and never change again. If a_i is his minmax action, staying at this action is the best alternative to following the equilibrium punishment. This would give him a payoff of w_i . So suppose a_i is not i 's minmax action. Let z_i denote i 's second best payoff when the others are trying to minmax him. In other words, letting a_i^* denote i 's minmax action,

$$z_i = \max_{a_i \neq a_i^*} \left[\min_{a_{\sim i} \in A_{\sim i}} u_i(a_i, a_{\sim i}) \right].$$

By assumption, there is a unique action vector which gives i a payoff of w_i , so it must be true that $w_i > z_i$. Given that ε is chosen to satisfy (2), $w_i - z_i > \varepsilon$. If i never changes actions again, his payoff is z_i at best. If he changes to his minmax action, his payoff is $w_i - \varepsilon$. Clearly, then, it is optimal for him to change to his minmax action. In short, i 's payoff if he does not follow the equilibrium punishment is w_i minus the switching cost if he is not already playing his minmax action, or $w_i - I_i(a_i)\varepsilon$. So i is indifferent between following the equilibrium punishment and not. In short, the target of a punishment has no incentive to deviate prior to the random determination of u .

To complete the proof that this is a subgame perfect equilibrium, consider any history for which there has been no deviation and any player i . If i does not deviate, his payoff will be at least u'_i (more if he has already carried out some changes of action). If i deviates, his expected payoff will be w_i at best. Since $u' \geq u \geq w$, i has no incentive to deviate.

Consider any history which puts us in punishment mode (i, a_i) and any $j \neq i$. Does j have an incentive to deviate prior to the realization of the public randomization determining u ? If j does not deviate, his payoff is at least

$$q_i(a_i)\underline{u}_j^i + [1 - q_i(a_i)]\bar{u}_j - 2\varepsilon.$$

(This would be the case if j has to switch actions at this point to punish the target and will have to switch again once u is realized.) If j deviates, his payoff will be w_j at best. Because $\underline{u}_j^i \geq w_j$, a sufficient condition for j to not deviate is

$$q_i(a_i)w_j + [1 - q_i(a_i)]\bar{u}_j - 2\varepsilon \geq w_j$$

or $[\bar{u}_j - w_j][1 - q_i(a_i)] \geq 2\varepsilon$. From the definition of $q_i(a_i)$ above, we see that a sufficient condition for this for $\varepsilon \leq 1$ is

$$[\bar{u}_j - w_j] \frac{k_i \varepsilon + w_i - \underline{u}_i^i - \varepsilon}{\bar{u}_i - \underline{u}_i^i + 1} \geq 2\varepsilon.$$

From the definition of k_i ,

$$k_i \varepsilon \geq \underline{u}_i^i - w_i + \varepsilon + 2\varepsilon \frac{\bar{u}_i - \underline{u}_i^i + 1}{\bar{u}_j - w_j}.$$

Hence a sufficient condition is

$$\frac{\bar{u}_j - w_j}{\bar{u}_i - \underline{u}_i^i + 1} 2\varepsilon \frac{\bar{u}_i - \underline{u}_i^i + 1}{\bar{u}_j - w_j} \geq 2\varepsilon,$$

which is obviously true. Hence no player, the target or otherwise, will deviate from a punishment prior to the realization of the public randomization.

Hence it only remains to show that no player will deviate after the realization of the randomization. Consider any player i who played a_i in the period before the realization. If i deviates from the equilibrium by playing \hat{a}_i , we move into an (i, \hat{a}_i) punishment mode and i 's payoff is $w_i - I_i(\hat{a}_i)\varepsilon$ minus ε if $\hat{a}_i \neq a_i$. If, instead, i follows the equilibrium, his payoff is u_i minus ε if he must switch actions, where u is the outcome of the random draw. Recall that $u_i \geq w_i$. First, suppose $u_i > w_i$. By (2), then, $u_i - \varepsilon > w_i$. Hence i has no incentive to deviate since the worst he could do by following the equilibrium is strictly better than the best he could do by deviating. Suppose, then, that $u_i = w_i$. Then i 's payoff from following the equilibrium is $w_i - I_i(a_i)\varepsilon$. If $\hat{a}_i = a_i$, this is exactly what i would get if he deviated. If $\hat{a}_i \neq a_i$, then i 's payoff to deviating is

$$-\varepsilon + w_i - I_i(\hat{a}_i)\varepsilon \leq w_i - \varepsilon \leq w_i - I_i(a_i)\varepsilon.$$

Hence either way, i has no incentive to deviate.

This demonstrates that $c(U_{\geq}) \cap W \subseteq \lim_{\varepsilon \downarrow 0} \mathcal{U}(\varepsilon, 1)$. To complete the proof, then, suppose $u \notin c(U_{\geq}) \cap W$. We now show that there is no equilibrium payoff nearby. Since $c(U_{\geq}) \cap W$ is closed and does not contain u , there is an $\varepsilon > 0$ such that for every u' within ε of u , $u' \notin c(U_{\geq}) \cap W$. Choose such an ε such that for all a and i with $u_i(a) < w_i$, we have $u_i(a) < w_i - \varepsilon$. Suppose, contrary to our claim, that there is a u' within ε of u with $u' \in \mathcal{U}(\varepsilon, 1)$. Obviously, if u' is not feasible, it cannot be an equilibrium payoff. So there must be some action profiles a^1, \dots, a^Z and strictly positive numbers $\alpha_1, \dots, \alpha^Z$ such that $\sum_z \alpha_z = 1$ and

$$u' \leq \sum_z \alpha_z u(a^z).$$

If $u' \not\geq w$, then it clearly cannot be an equilibrium payoff vector since each player i can certainly guarantee himself w_i . Hence $u' \geq w$.

Since $u' \notin c(U_{\geq}) \cap W$, then, it must be true that at least one of the a^1, \dots, a^Z profiles has $u_i(a^z) < w_i$ for some i . By our assumption on ε , this implies that at least one of the a^1, \dots, a^Z profiles has $u_i(a^z) < w_i - \varepsilon$. This does not necessarily imply that this particular a^z is played in equilibrium. The decomposition of a payoff into a convex combination of the other payoffs is not unique in general, so u' could be generated without a^z ever played. On the other hand, it is clear that there is some a^z with $u_i(a^z) < w_i$ which is played infinitely often with strictly positive probability. Otherwise, no such a^z could receive any weight in the convex combination, meaning that $u' \in c(U_{\geq}) \cap W$, a contradiction. So let \hat{a} denote any such profile and let i denote any player for whom $u_i(\hat{a}) < w_i$ (and hence $u_i(\hat{a}) < w_i - \varepsilon$). By hypothesis, there is a strictly positive probability that \hat{a} is played infinitely often in equilibrium.

Let \mathcal{P} denote the set of paths (infinite sequences of action profiles) in the support of the equilibrium. Let μ denote the probability distribution over \mathcal{P} induced by the equilibrium (including the effect of the public randomizations if strategies are based on these). Let $\mathcal{P}^*(a)$ denote the set of paths in \mathcal{P} which eventually absorb at action profile a — that is,

$$\mathcal{P}^*(a) = \{(a^1, a^2, \dots) \in \mathcal{P} \mid \exists T \text{ such that } a^t = a, \forall t \geq T\}.$$

Similarly, let \mathcal{P}_d denote the set of paths in \mathcal{P} which do not absorb — that is, $\mathcal{P}_d = \mathcal{P} \setminus \cup_{a \in A} \mathcal{P}^*(a)$. It is not hard to see that $\mu(\mathcal{P}_d) = 0$. If this were not zero, then the expected total switching costs of the players would necessarily be infinite, meaning that some player's payoff is $-\infty$, so obviously his strategy cannot be optimal.

Our selection of \hat{a} implies that $\mu(\mathcal{P}^*(\hat{a})) > 0$. Recall that we chose \hat{a} to be some profile played infinitely often with strictly positive probability. Since no path in $\mathcal{P}^*(a)$ for $a \neq \hat{a}$ has this property and since $\sum_{a \in A} \mu(\mathcal{P}^*(a)) = 1$, we must have $\mu(\mathcal{P}^*(\hat{a})) > 0$.

Let $\mathcal{P}^t(a)$ denote the set of paths in \mathcal{P} with $a^t = a$. For any path $p \in \mathcal{P}^*(a)$, $a \neq \hat{a}$, there is a T such that $p \notin \mathcal{P}^t(\hat{a})$ for all $t \geq T$. That is, if a path eventually stays at $a \neq \hat{a}$ forever, it must visit \hat{a} for a last time at some finite date. Hence for all $a \neq \hat{a}$,

$$\mathcal{P}^*(a) \cap \mathcal{P}^t(\hat{a}) \rightarrow \emptyset$$

as $t \rightarrow \infty$. On the other hand, consider any $p \in \mathcal{P}^*(\hat{a})$. By definition, there is a T such that this path has $a^t = \hat{a}$ for all $t \geq T$. Hence there is a T such that this path is in $\mathcal{P}^t(\hat{a})$ for all $t \geq T$. Hence

$$\mathcal{P}^*(\hat{a}) \cap \mathcal{P}^t(\hat{a}) \rightarrow \mathcal{P}^*(\hat{a})$$

as $t \rightarrow \infty$.

In light of this, consider

$$\mu(\mathcal{P}^*(\hat{a}) \mid \mathcal{P}^t(\hat{a})) = \frac{\mu(\mathcal{P}^*(\hat{a}) \cap \mathcal{P}^t(\hat{a}))}{\mu(\mathcal{P}^t(\hat{a}))}.$$

Since all the $\mathcal{P}^*(a)$ sets are disjoint and their union has probability 1, we can rewrite this as

$$\frac{\mu(\mathcal{P}^*(\hat{a}) \cap \mathcal{P}^t(\hat{a}))}{\sum_{a \in A} \mu(\mathcal{P}^*(a) \cap \mathcal{P}^t(\hat{a}))}.$$

Clearly, as $t \rightarrow \infty$, this converges to

$$\frac{\mu(\mathcal{P}^*(\hat{a}))}{\mu(\mathcal{P}^*(\hat{a}))} = 1.$$

(Note that this is well defined since $\mu(\mathcal{P}^*(\hat{a})) > 0$.) In short, $\mu(\mathcal{P}^*(\hat{a}) \mid \mathcal{P}^t(\hat{a})) \rightarrow 1$ as $t \rightarrow \infty$.

Consider player i , the player for whom $u_i(\hat{a}) < w_i$. Fix some t and consider the following strategy for player i : follow the equilibrium strategy until the equilibrium strategies call for \hat{a} to be played at period t . Then deviate to the pure minmax action forever after. Clearly, since this is an equilibrium, this alternative strategy cannot be better for i for any choice of t . In comparing i 's payoff in the equilibrium to i 's payoff from the alternative strategy, obviously, we can condition on the set of paths for which \hat{a} is played at time t — for any other paths, the payoff difference is zero. If player i deviates at time t as specified, his expected payoff from that point onward is at least $w_i - \varepsilon$. But as $t \rightarrow \infty$, player i 's expected continuation payoff if he does not deviate is converging to $u_i(\hat{a})$. Since $u_i(\hat{a}) < w_i - \varepsilon$, there is some large t for which i strictly prefers the deviation, a contradiction. ■

References

- [1] D. Abreu, P. Dutta, and L. Smith, “The Folk Theorem for Repeated Games: A NEU Condition,” *Econometrica*, **62**, July 1994, 939–948.
- [2] S. Chakrabarti, “Characterizations of the Equilibrium Payoffs of Inertia Supergames,” *Journal of Economic Theory*, **51**, 1990, 171–83.
- [3] P. Dutta, “A Folk Theorem for Stochastic Games,” *Journal of Economic Theory* **66** (1995), 1–32.
- [4] P. Dutta, “Collusion, Discounting, and Dynamic Games,” *Journal of Economic Theory* **66** (1995), 289–306.
- [5] D. Fudenberg and E. Maskin, “On the Dispensability of Public Randomization in Discounted Repeated Games,” *Journal of Economic Theory*, **53**, April 1991, 428–438.
- [6] D. Fudenberg and J. Tirole, *Game Theory*, Cambridge: MIT Press, 1991.
- [7] B. Lipman and R. Wang, “Switching Costs in Frequently Repeated Games,” *Journal of Economic Theory*, **93**, August 2000, 149–190.
- [8] Q. Wen, “The ‘Folk Theorem’ for Repeated Games with Complete Information,” *Econometrica*, **62**, July 1994, 949–954.