Forward Induction and Public Randomization†

Faruk Gul
Northwestern University

and

David G. Pearce
Yale University

January 1992
Revised June 1996

† We thank Ennio Stacchetti for his help and Dilip Abreu for discussions that led us to this project. Both authors gratefully acknowledge financial support from the Alfred P. Sloan Foundation and the National Science Foundation.

Forward Induction

Faruk Gul
Department of Economics
Princeton University
Princeton, NJ 08544

Abstract

We show that if the set of subgame perfect equilibrium payoffs of a stage game is convexified by the introduction of public random variables, then forward induction loses all its power as an argument for refining Nash equilibrium. We establish that our argument does not depend on a convenient choice of off-equilibrium path beliefs.

JEL Classification C72

1. Introduction

The purpose of this paper is to suggest that forward induction is much less powerful than is usually supposed as a tool for refining noncooperative solution concepts. In any finite extensive form game having a "stage game" structure, the logic of forward induction is so fragile that it cannot survive the introduction of public randomization. Since publicly observed random variables are available to players in most settings of interest, perhaps a reconsideration of the status of forward induction is warranted.

Following its introduction by Kohlberg and Mertens [7], the term "forward induction" has taken on many shades of meaning. Here we use it to refer to the process by which players infer from a choice of some player what he or she is likely to do subsequently. This is distinct from phenomena such as signaling private information about one's type (Cho and Kreps [3]), but is of the essence for understanding the signaling of strategic intent. Taking account of the possibility of signaling strategic intent is important not just for equilibrium analysis but also for studying rationalizable strategic behavior in the extensive form. Our principal focus, however, will be on the force of forward induction as an equilibrium refinement.

Although there is a substantial and stimulating literature on forward induction (see especially Kohlberg and Mertens [7], Osborne [10] and van Damme [12]), there is no consensus on what the term "forward induction" should mean. Fortunately, the point of this paper can be made without resolving that issue. We show that even in its strongest form, forward induction is impotent when the set of equilibrium payoffs in each subgame is convexified via public randomization.

In section 2 we present a well-known example to illustrate our results. Section 3 describes finite stage games and defines for any stage game Γ an associated game Γ^c with public randomizing devices at the beginning of each period. Section 4 gives two definitions, one stronger than the other, of forward induction-proof equilibrium. We then present two propositions concerning the solution concepts. The first proposition, relatively transparent, shows how to take *any* subgame perfect equilibrium σ of a game Γ and construct another subgame perfect equilibrium σ^f of Γ^c having the same equilibrium path and satisfying forward induction-proofness. For the purposes of this proposition, we use the weaker notion of forward induction-proofness. One could object that no convincing justification is given for the out-of-equilibrium beliefs needed to support the equilibrium σ^f; if the beliefs seem contrived, then the force of the proposition as an assault on forward induction is diminished. Also, one might insist on a stronger notion of forward induction-proofness.

The second proposition is offered as an answer to both of the preceding objections. Given an equilibrium σ, partition the actions of each player at each stage into a relevant set and an irrelevant set as follows: an irrelevant action a_i for player i is one such that it is (strictly) unprofitable for i to deviate from σ by using that action no matter which subgame perfect solutions were to govern play following the deviation (i.e., in subgames that are reached by action profiles (a_i, a_{-i}) for a_{-i} that are played with positive probability in σ). In short, no one could conceivably gain by using an irrelevant action to signal strategic intent. The second proposition says that for a generic class of stage games, any Γ in that class, any subgame perfect equilibrium σ of Γ, and any $\epsilon > 0$, there exists a subgame perfect equilibrium σ^g of Γ^c within ϵ of σ that is strongly forward induction-proof (that is, uses *all* relevant actions). No role is played by expectations held in out-of-equilibrium contingencies and hence no opportunity is left for forward induction to rule out any subgame perfect equilibrium. In section 5, we discuss the interpretations of public randomization devices and our results.

2. An Example

Consider the extensive form game in figure 1.[1] It has two subgame perfect outcomes: $(7, 7)$ and $(10, 2)$. To support the $(7, 7)$ outcome, player 2 when reached must choose f with sufficient probability. Consider any equilibrium that supports the $(7, 7)$ outcome. Suppose player 1 deviates and player 2's information set is reached. Player 2 should understand that this means player 1 has forgone a payoff of 7. If player 1 had played d, she could not have achieved a payoff as high as 7 regardless of what player 2 did. Thus, player 2 must conclude that player 1 must have chosen c and should respond with e. This yields player 1 a payoff of 10. Since player 1 can anticipate player 2's line of reasoning, she should deviate

[1] This game is closely related to games studied by Kohlberg and Mertens [7] and van Damme [12]. Kohlberg and Mertens combine the two information sets of player 1. Hence, their version violates our definition of a stage game (see Definition 1 below).

picture

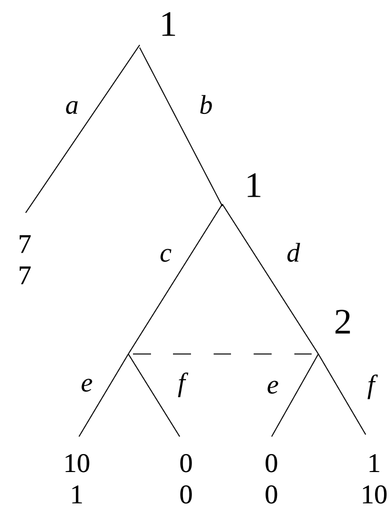

Figure 1.

from a, her action in the proposed equilibrium. Thus, no equilibrium in which player 1 plays a is forward-induction proof.

Next, imagine a publicly observable random variable θ whose realization will be observable by both players after the time at which player 1 has to choose between a and b and before any further choices are made. Suppose θ has a continuous distribution. Let G be the cumulative distribution of θ, and let θ^* satisfy $G(\theta^*) = \frac{2}{3}$. If player 1 believes that player 2 will choose e whenever $\theta \leq \theta^*$ and f otherwise, then player 1 will be indifferent between choosing a or choosing b followed by c if $\theta \leq \theta^*$ and by d otherwise. To see this, note that the latter strategy yields an *ex ante* payoff of $\frac{2}{3} \times 10 + \frac{1}{3} \times 1 = 7$, which is the same as the payoff associated with choosing a. Consider the subgame perfect equilibrium of the convexified game above in which player 1 chooses a at her first information set and in the subgame starting with nature's choice of θ, player 1 and player 2 behave as described above. Now if player 2 is reached, he concludes that player 1 has failed to choose a, not due to an intent to signal, but out of indifference. At the time at which player 1 chooses to deviate from the equilibrium-prescribed action a, rationalizing her decision to do so does not necessitate an argument that precludes player 1 from ever choosing d. It simply requires that d not be played too often. The proposed equilibrium (of the convexified game) above ensures that d is not played too often. Our Proposition 1 generalizes the preceding analysis to arbitrary stage games. In section 4 (Proposition 1), we will show that any subgame perfect equilibrium can be rendered forward induction-proof as described in the above example in the presence of publicly observable random variables.

One could object to the beliefs assigned to the players after player 1's choice of b. Perhaps we should have required that, after the deviation by player 1, player 2 should seek an explanation that makes player 1 strictly better off compared to the equilibrium. Or perhaps the beliefs specified are inappropriate for some other reason.

Suppose we now adjust the original equilibrium so that player 1 chooses a with probability $1-\epsilon$ and b with probability ϵ. Then if b is chosen, player 1 chooses c and player 2 chooses e if $\theta \leq \theta^*$ and player 1 chooses d and player 2 chooses f if $\theta > \theta^*$. Note that this is a subgame perfect equilibrium of the convexified game in which there are no unreached subgames. Hence this equilibrium is immune from any forward induction argument. Moreover, by choosing ϵ small enough, the equilibrium distribution over terminal nodes can be made arbitrarily close to the outcome of the equilibrium that was eliminated in the non-convexified game. Our Proposition 2 generalizes this analysis to arbitrary stage games for generic payoffs. The idea of Proposition 1 is essentially as straightforward as its application to the example above. Proposition 2 is more delicate because an ϵ perturbation will affect payoffs in the previous stage; this complicates the analysis when more than one player has a non-trivial choice at a given stage.

3. Definitions

This paper considers finite extensive form games with perfect recall (see Kuhn [8] or Selten [11]). Our results concern games having a stage game structure. Such a game can be thought of as taking place over T periods, with decision-makers in any period t fully informed about what happened in the first $t-1$ periods but totally ignorant about others' play in period t.

Definition 1: *For any positive integer T, a finite extensive game Γ with perfect recall is a stage game of length T if there exists a surjective map f from the choice nodes of Γ to the set $\{1, \ldots, T\}$ such that* *(i)* *for each* $t \in \{1, \dots, T\}$ *and each node* x *such that* $f(x) \geq t$, x *belongs to some subgame whose origin* y *satisfies* $f(y) = t$; *and*

(ii) *if subgame* τ *has origin* z *with* $f(z) = t$, *then if an information set* I *for player* i *in* τ *contains an* x *with* $f(x) = t$, *any choice node* x' *of* i *in* τ *such that* $f(x') = t$ *is also in* I.

Condition (i) ensures that a "true" subgame follows every $(t-1)$-period history. Thus, what happened in the first $t - 1$ periods can be regarded as common knowledge at the beginning of period t. Condition (ii) requires that when making a choice in period t, no player has any information concerning other players' choices in period t. Consequently play within a period is, in the informational sense, simultaneous. Notice that not all players are necessarily active in a particular period t following some history h. A game of perfect information, for example, is a stage game. Moreover, the "component game" played in period t may depend on the history h, which may in turn include chance moves by nature. Thus, finite dynamic and stochastic games are included in this class.

Definition 2: *Let* $\theta_1, \dots, \theta_T$ *be independent random variables, each uniformly distributed on* $[0, 1]$. *Modify the stage game* Γ *of length* T *as follows: at the beginning of each period* t, *before that period's play commences, the realization of the random variable* θ_t *is publicly observed. Thus, choices in period* t *can be conditioned on* θ_t *(in addition to the* $(t-1)$*-period history). We call the new game the* convexification *of* Γ *and denote it* Γ^c.

The particular choice of distribution for θ_t is not important; any non-degenerate continuous distribution will suffice. The role of the random variables is to convexify the set of subgame perfect equilibrium values from the beginning of each period onward. The opportunity of conditioning on the realizations of a continuous random variable substantially complicates the strategy sets of the players and introduces a number of technical issues. In order that expected payoffs may be computed, strategies must satisfy obvious measurability requirements. Moreover, it is possible that players' strategies at stage t could depend on the realization of earlier random variables. Our main objective is to show that if the SPE payoff set is convex, forward induction loses all its force. For this it will be sufficient to consider equilibria in which strategies do not depend on earlier realizations. Furthermore, the simplicity of the strategies we construct will make measurability easy to verify.

4. Forward Induction in Convexified Games

Forward induction usually involves an inference that, because some player i has chosen a particular action or sequence of actions, there are certain things i will not do subsequently in the game. Some applications require only iterative dominance reasoning, while others use a particular equilibrium as a *status quo* point of reference. In some cases, the aforementioned "inference" is based on lines of reasoning that are less than compelling. We do not attempt a sophisticated formulation of forward induction here. This section presents two crude definitions of forward induction-proof equilibrium, both of which are biased toward eliminating candidate solutions. We then give two results showing that once public randomization is allowed for, our first definition of forward induction admits all subgame perfect equilibrium paths as solutions. Moreover, even our more restrictive definition of forward induction permits equilibrium distributions over terminal nodes arbitrarily close to any subgame perfect equilibrium distribution for generic stage games.

Standard principles of backward induction are incorporated into the definition by recursion on T, the length of the game.

Definition 3: *An action a_i at stage t of a stage game (or a convexified stage game) is called threatening relative to σ if*

(i) a_i is strictly suboptimal; and

(ii) there exists another SPE γ of Γ (or the convexification of Γ) such that if i (after observing θ_t) deviates to a_i and all others play according to σ in period t and γ in the remaining $T-t$ periods, then the deviation is strictly profitable for i.

To define a *relevant* action, we replace (i) above with "a_i is not used," and we replace the phrase "strictly profitable" in (ii) above with "not strictly unprofitable."

Clearly, every threatening action is a relevant action.

Definition 4: *Let Γ be a stage game of length T, or a convexification of such a game. A forward induction-proof equilibrium of Γ is defined recursively as follows:*

(1) for $T = 1$, a strategy profile σ of Γ is forward induction-proof if σ is a Nash equilibrium of Γ;

(2) for $T = 2, 3, \ldots$, a profile σ of Γ is forward induction-proof if it is a subgame perfect equilibrium of Γ and if

(a) the continuation equilibria induced by σ on the $(T-1)$-period subgames following the first period are forward induction-proof; and

(b) no player has a threatening action (relative to σ).

A *strongly forward induction-proof equilibrium* is defined by replacing "threatening" with "relevant."

Suppose that in some subgame perfect equilibrium (hereafter SPE), player i did have a threatening action a_i. The idea is that if i were to deviate by playing a_i, others would conclude that i did not believe that σ would henceforth govern play. Players might then coordinate on some other equilibrium lucrative enough for i in the remaining periods to explain the deviation to a_i.

Note that in defining a threatening action (i.e., for forward induction-proofness) we considered only those deviations that yield a payoff strictly below the equilibrium payoff of player i, given the behavior prescribed by the equilibrium σ for i's opponents. According to this definition, an action which yields the same payoff as the equilibrium payoff for player i provides no opportunity for strategic signaling. Such an action yields a payoff no better or worse than what player i would have obtained in equilibrium and hence does not constitute an implicit speech in the sense of Cho and Kreps [3].

On the other hand, for defining a strongly forward induction-proof equilibrium, any unused action is viewed as a possible vehicle for signaling. Moreover, we require that following a deviation to such an action, every SPE yield a *strictly* lower payoff. We offer the following observations about the two definitions of forward induction-proofness presented above:

(1) In defining both notions of forward induction-proofness, we do not insist that the alternative equilibrium on which the players might coordinate after player i deviates be forward induction-proof. This makes it easier to find threatening or relevant actions, and renders our definitions more restrictive.

(2) For the reason described in item (1) above, it is easy to verify that a strongly forward induction-proof equilibrium is also a forward induction-proof equilibrium since a threatening action is always a relevant action.

(3) The notion of a strongly forward induction-proof equilibrium is the strongest notion of signaling strategic intent by deviating from equilibrium that is consistent with subgame perfection. If a player has no threatening action in the sense of Definition 4, then he will have a strict incentive not to deviate provided he believes that every deviation will be followed by a subgame perfect equilibrium in the subsequent subgame.

(4) Both definitions yield the unique equilibrium identified by maximal elimination of iterative dominated strategies in Ben-Porath and Dekel [1] as the only forward induction-proof equilibrium. Of course, forward induction-proofness as defined above is not a more stringent requirement than iterative dominance in all stage games since iterative elimination can remove equilibria even in single-stage games, whereas for such games forward induction-proofness has no bite. Similarly, our notions of forward induction-proofness identify the "right equilibrium" in the example discussed in section 2 above and in the two examples considered by van Damme [12]. Forward induction in our sense is implied by the notion of forward induction presented in van Damme [12]. However, it is difficult to compare our notion to the "never a weak best response" property (NWBR) of Kohlberg and Mertens (see Proposition 6.B of their paper). This is due to the fact that Kohlberg and Mertens [7] define forward induction in terms of the normal form and as a property of a component of equilibrium.[2]

The comparison between our notion of forward induction and the intuitive criterion of Cho and Kreps [3] or Cho [2] is somewhat easier but uninstructive. These refinements do not eliminate any SPE of stage games, even without convexification. The Cho-Kreps intuitive criterion is relevant only for signaling games. Cho's [2] refinement eliminates sequential equilibria by considering one-shot deviations and restricting beliefs at information sets that are reached by such deviations but not reached in the

[2] The only result (or definition) that Kohlberg and Mertens offer regarding forward induction is Proposition 6.B: "Every stable set contains a stable set of the game obtained by deleting a strategy which is an inferior response to all the equilibria of the set." Hence, Kohlberg and Mertens offer no guidance to anyone interested in forward induction proofness as an extensive form solution concept.

original equilibrium (even after conditioning on the information set at which the deviation is to occur). The refinement compares, conditional on reaching the information set at which the deviation occurred, what the deviator would receive in the purported equilibrium versus what he receives with the deviation given a reasonable response by the remaining players. Deviations that can never improve the payoff of the deviator are called BAD. It is required that observers of deviations assign zero probability to BAD deviations when good deviations are possible. In a stage game, conditional on an information set h being reached, any information set that contains a successor of h is either degenerate or is reached with probability 1. In either case there is no room for Cho's refinement to eliminate equilibria. Thus our notions of forward induction are, by default, more restrictive.

picture

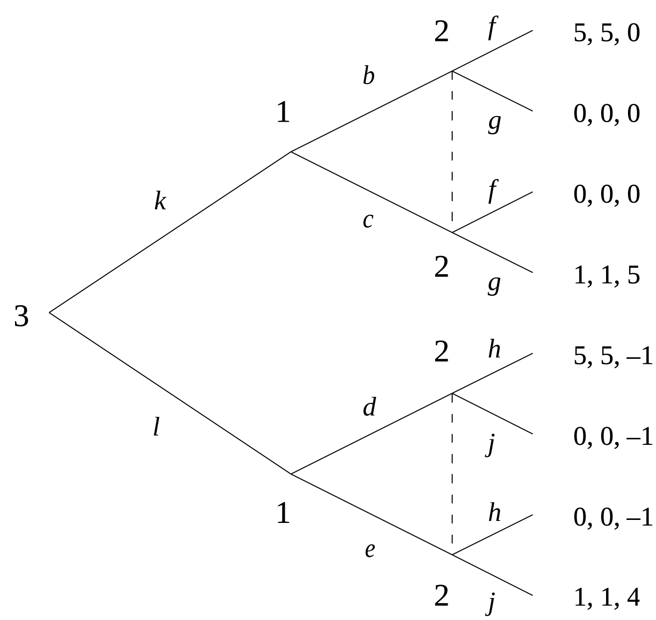

Figure 2.

For understanding why even the weaker of our two notions of forward induction is too restrictive, the game in figure 2 is particularly telling. It seems reasonable to argue that in whichever subgame they are reached in period 2, players 1 and 2 will play their unanimously preferred equilibrium and each receive 5. Player 3 should therefore choose action k (choosing l is like "burning a dollar" *a la* Ben-Porath and Dekel [1] and van Damme [12]). According to our definition, this equilibrium is upset by the threatening action l, because although l is a bad move for player 3 in the candidate equilibrium, it would be a good move if it convinced players 1 and 2 to switch to the equilibrium in which they both get a payoff of 1. But does this make any sense? Player 3 plays no strategic role in either second-period subgame, so there is no scope for player 3 to signal strategic intent. Perhaps player 3 can signal her beliefs; but why should players 1 and 2 care? Admittedly, it is conceivable that they could be influenced by player 3's choice—perhaps each of them thinks that the other thinks the way player 3 does, and player 3 apparently thinks that playing l will result in the outcome $(1, 1, 4)$. But this is a slim argument for throwing out the original equilibrium. For this and other reasons, it seems fair to say that the definition rejects candidate solutions too readily. In spite of these restrictive notions of forward induction, Propositions 1 and 2 show that in the presence of public randomization the definition imposes essentially no restriction beyond subgame perfection.

Since the strategy sets in the two games Γ and Γ^c are different, the statement that an equilibrium of one is the same as or close to an equilibrium of the other requires some clarification. We will refer to the finite set of terminal nodes of Γ as physical outcomes. Thus, each profile of mixed strategies in Γ yields probability distributions over physical outcomes. Similarly, the more complicated strategy profiles of Γ^c can also be associated with probability distributions over physical outcomes of Γ. We will say that a strategy profile σ^c of Γ^c is close to σ of Γ if the distribution over physical outcomes implied by σ^c is close to the distribution over physical outcomes implied by σ.

Proposition 1: *Let Γ be a stage game of length T and σ be any SPE (subgame perfect equilibrium) of Γ^c, the convexification of Γ. There exists a forward induction-proof equilibrium σ^f of Γ^c having the same equilibrium distribution over physical outcomes as σ.*

Proof: It is easy to check that the presence of the public randomizing devices in Γ^c ensures the convexity of the subgame perfect equilibrium value set for the subgame following any t-period history. We establish the main proposition via an induction on T. It holds trivially for $T = 1$. Suppose that the proposition has been established for $T = 1, 2, \ldots, t-1$. We need to show it holds for t.

Let σ be an SPE of Γ^c, a convexified stage game of length t. For each action a_i for player i that is threatening in the first period (with respect to σ), there is by definition some collection $\succ(a_i)$ of *preferable* $(t-1)$-period subgame perfect equilibria $\succ(a_{-i}, a_i)$ (for each a_{-i} played with positive probability in σ) such that in the profile agreeing with $\succ(a_{-i}, a_i)$ in subgames following the choice a_i, and with σ everywhere else, i would strictly gain by a "one-shot deviation" to a_i. Thus, each $\succ(a_{-i}, a_i)$ is an equilibrium of the subgame of Γ^c starting with the initial node reached by (a_{-i}, a_i), while $\succ(a_i)$ refers to the collection $\succ(a_{-i}, a_i)$ (one for each a_{-i} played with positive probability). Notice that depending on which subgame perfect continuation we choose (one from σ or $\succ(a_i)$), a_i is either strictly undesirable or strictly desirable, relative to the first-stage behavior for i stipulated by σ. By the convexity of the set of SPE values after any first-period history (noted earlier), we can therefore choose a collection of SPE's $\sim(a_i)$ for the subgames in question such that in the profile agreeing with $\sim(a_i)$ in those subgames and with σ elsewhere, i is indifferent between a_i and the behavior stipulated by σ. Choose $\sim(a_i)$ to be forward induction-proof in every $(t-1)$ subgame (this is possible by the inductive hypothesis and by the axiom of choice, which is needed because θ_t takes an uncountable number of values).

Define a new strategy profile σ^f in Γ^c as follows:

(1) In period 1, σ^f agrees with σ.

(2) Following any first-period history in which the number of players who have chosen actions that are threatening with respect to σ in Γ^c is *not* exactly one, replace the continuation equilibrium induced by σ on the subgame by a forward induction-proof equilibrium having the same distribution over physical outcomes. That this is possible follows from the inductive hypothesis.

(3) Following any first-period history in which exactly one player i has chosen a threatening action (with respect to σ in Γ^c), say a_i, replace the continuation equilibrium induced by σ in the subsequent $(t-1)$-period subgame by $\sim(a_i)$, defined above.

We claim that σ^f is forward induction-proof. It is an SPE because

(i) by construction its continuation profiles following all first-period histories are subgame perfect; and

(ii) no deviation in the first period is profitable: payoffs to non-threatening actions are unchanged, and for each action a_i that is threatening, $\sim(a_i)$ was chosen so that in σ^f the relevant player is indifferent between a_i and the equilibrium choice.

The indifference noted in the preceding sentence means that in the first period, there are no threatening actions with respect to σ^f in Γ^c. This, together with the fact that all continuation profiles following period one histories were chosen to be forward induction-proof, implies that σ^f is a forward induction-proof equilibrium. ∎

The import of Proposition 1 is that forward induction is powerless to refine the set of subgame perfect equilibria of stage games with public randomizations. One might question the seriousness of this result by saying that, in the equilibrium σ^f defined in the proof, the beliefs off the equilibrium path concerning future play, particularly following actions that were threatening with respect to σ, are contrived. Perhaps the construction of σ^f unduly exploits the freedom to manipulate these expectations at unreached information sets.

Moreover, the definition of forward induction requires that if a deviation does not yield a strictly lower payoff than the equilibrium in question, then it can not be used as a "credible" implicit speech. In a forward induction-proof equilibrium, a player may interpret player i's deviation as an attempt to secure this equilibrium payoff in a different manner, even when it is possible to interpret it as an attempt to attain a strictly higher payoff.

Proposition 2 is offered to give assurance that the result is not essentially dependent on a convenient choice of off-equilibrium path conjectures, nor on a heavy-handed use of the deviating player's indifference. Starting with an SPE σ of Γ it constructs a nearby SPE σ^g of Γ^c that is forward induction-proof and that uses all the previously relevant actions with positive probability in equilibrium. In σ^g, the only unused actions available to players as "signals" are those that are "irrelevant" in the sense of Definition 3: they cannot be profitable whatever equilibrium they trigger in the remainder game.

Hence Proposition 2 establishes that given convexification any extensive form refinement that builds on subgame perfection and operates by restricting beliefs on unreached information sets is impotent in generic stage games. This follows from the fact that we can render all deviations irrelevant. Player i is always worse off by choosing a_i in period t given what i's opponents are doing in period t and any subgame perfect continuation after period t.

Proposition 2: *Let Γ be a stage game with ℓ terminal nodes. There exists an open set $U^* \subset I\!R^{n\ell}$ such that $I\!R^{n\ell} \backslash U^*$ has Lebesgue measure zero and, for any $u \in U^*$, SPE σ of $\Gamma(u)$, and $\epsilon > 0$ there exists a strongly forward induction-proof equilibrium σ^g of $\Gamma^c(u)$ such that the probabilities of any given physical outcome when play is governed by σ^g and σ, respectively, differ by less than ϵ.*

The proof of Proposition 2 is given in the appendix.

picture

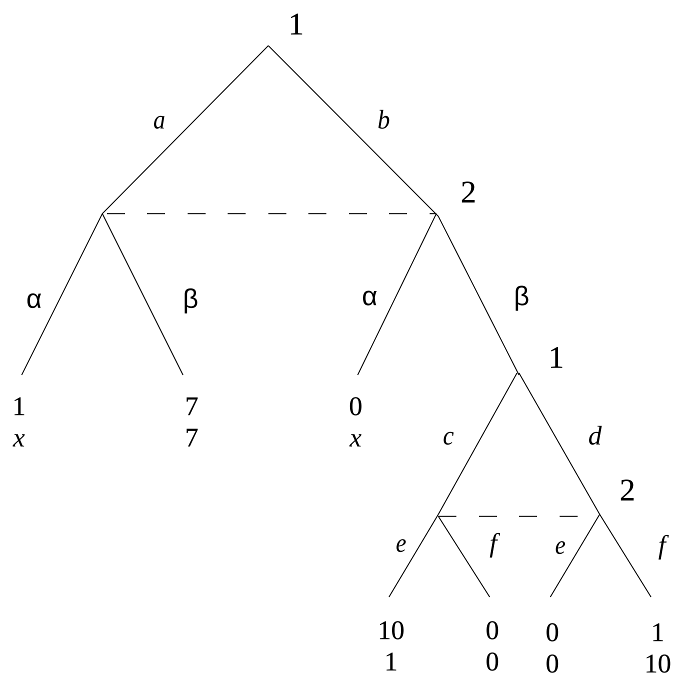

Figure 3.

To see how the proof works and why the genericity condition is needed, consider the slightly modified version of figure 1 contained in figure 3. Note that (a, β) is a subgame perfect equilibrium outcome of this game whenever $x \leq 7$. Moreover the analysis of the deviation by player 1 to the strategy b in stage 1 is analogous to the analysis of the same deviation in figure 1. To construct an equilibrium of the convexified game in which player 1 is deprived of the signaling opportunity that this deviation affords, we would once again like to construct an equilibrium in which b is used with some small probability, ϵ. This necessitates that we construct a continuation payoff for player 1 after action b that yields her a payoff of 7. Consider using the same strategies for both players as described in section 2.

This will yield a payoff of two to player 1 and a payoff of $(1-\epsilon)7+\epsilon\big(\frac{2}{3}\times 10+\frac{1}{3}\times 1\big)$ to player 2. If $x<7$, we are done. We can choose ϵ as small as we wish and have an equilibrium path that is arbitrarily close to (a,β). Moreover no player will have an unused strategy that could be used as a signal to unravel the equilibrium. Note, however, that any strategy profile in the subgame that is reached with (b,β) in the convexified game yields a total payoff no greater than 11 for the two players. Thus any subgame continuation that yields 7 to player 1 must yield less than 7 to player 2. Hence, if $x=7$ then we can not construct a SPE in which player 1 plays a with a high probability, b with some positive probability, and player 2 plays β with any probability. This is due to the fact that if player 2 is achieving a payoff close to 7 and is indifferent between a and b, then player 2 must be getting less than 7 after playing β (while he can guarantee 7 with α). Thus, for $x=7$ the payoff vector of Γ is not in U^*. For $x=7$ there exists no σ^g, an equilibrium of the convexified game, such that for the SPE $\sigma=\big((a,d),(\beta,f)\big)$ conditions 2a and 2b in the definition of strongly forward induction-proof are satisfied.

In order to be able to construct an equilibrium σ^g close to the equilibrium σ of the original game we need to ensure that for every subgame perfect equilibrium of any subgame of Γ, there exists a neighborhood of the original payoff vector such that the SPE correspondence is continuous in the payoff vector within the neighborhood. The proof of genericity in the appendix establishes that this is the case for an open, full-measure set of payoffs.

5. Discussion

Convexifying an extensive form game by adding publicly observable random variables typically alters the set of equilibrium distributions over terminal nodes. However, the effect of convexification is sensitive to the solution concept used. For example, the Nash equilibrium distributions of Γ^c can be identified with the convex hull of the Nash equilibrium distributions of Γ. Proposition 1 above establishes that the forward induction-proof equilibrium distributions of Γ^c constitute a much larger set than the forward induction-proof distribution of Γ. Convexification has an important role because the game Γ^c is strategically very different from the game Γ. To see this, consider again the game in figure 1. For this game any strategy that reaches player 1's second information set and entails choosing action d at that information set is strictly dominated. This is not true for the convexified version of the game. As can be seen from our analysis of section 2, one can construct many equilibria in the convexified game in which player 1's second information set is reached and d is chosen. The fact that convexification saves certain actions from being dominated suggests that other refinements also lose their power to restrict equilibria in convexified games. Thus it seems clear that convexification of stage games in the manner described in this paper is very effective in rendering a refinement powerless. Our final task will be to discuss how such convexification can be interpreted or justified.

Our view of publicly observable random variables is similar to the standard view of mixed strategies. We consider them not as the outcome of pregame communication or design but rather a ubiquitous element of the environment. Such random variables are ignored in the formal description of the game the same way that private randomization (i.e., mixed strategies) is ignored. In our view, coordinating on a possibly payoff-irrelevant public random variable is entirely compatible with the non-cooperative spirit of Nash equilibrium, as is coordinating on a payoff-irrelevant history in the celebrated folk theorems of supergames. Moreover, the analysis of Proposition 2 establishes that one could, for generic stage games, perturb the payoffs so that the random variables on which agents condition are no longer payoff irrelevant and thus extend Harsanyi's [6] well-known purification result to our context.

In any event, our case for the impotence of forward induction goes beyond games where public randomizing devices play a role. More generally, we are noting that forward induction derives its power entirely from non-convexities in value sets (of equilibria in subgames). These sets may be convex for many reasons other than public randomization: think of an infinitely repeated game with little discounting, for example; or games with unrestricted simultaneous "cheap-talk" possibilities between stages.[3]

We do not, however, wish to suggest that Proposition 2 forces one to view all subgame perfect equilibrium as being on equal footing. The stage game structure, and hence the opportunity to observe a public random variable at the right time, is crucial for our analysis. Moreover, in certain games there might be other reasons to eliminate equilibria, perhaps because the prescribed behavior or beliefs *along* the equilibrium path are viewed as being implausible. We suggest only that the class of games in which the logic of forward induction alone leads to severe restriction is rather limited.

[3] Myerson [9] page 294-5, offers an analysis of the effect of communication through a mediator, similar to our discussion of convexification and dominance above.

6. Appendix

Genericity: The purpose of this section is to provide a set $U^* \subset I\!R^{\ell n}$ where n is the number of players and ℓ is the number of terminal nodes in Γ, such that for every $u \in U^*$, Proposition 2 applies for the game $\Gamma(u)$.

The notation needed for the subsequent discussion of genericity is prohibitively difficult for the case of an arbitrary Γ. Therefore, we will restrict attention to the simple two-stage 2-person game Γ depicted in figure 4. The game Γ is special for a number of reasons: first, every player is active in every stage, after any history; second, every player has exactly two action choices at every one of his information sets; third, there are two players; fourth, there are only two stages; and fifth, there are no chance nodes. The fact that none of the discussion below rests on the first three properties of Γ will be immediately obvious. There is a simple argument that shows how the analysis below can be extended to games with a larger number of stages and chance nodes. This we will present at the end of this section.

picture

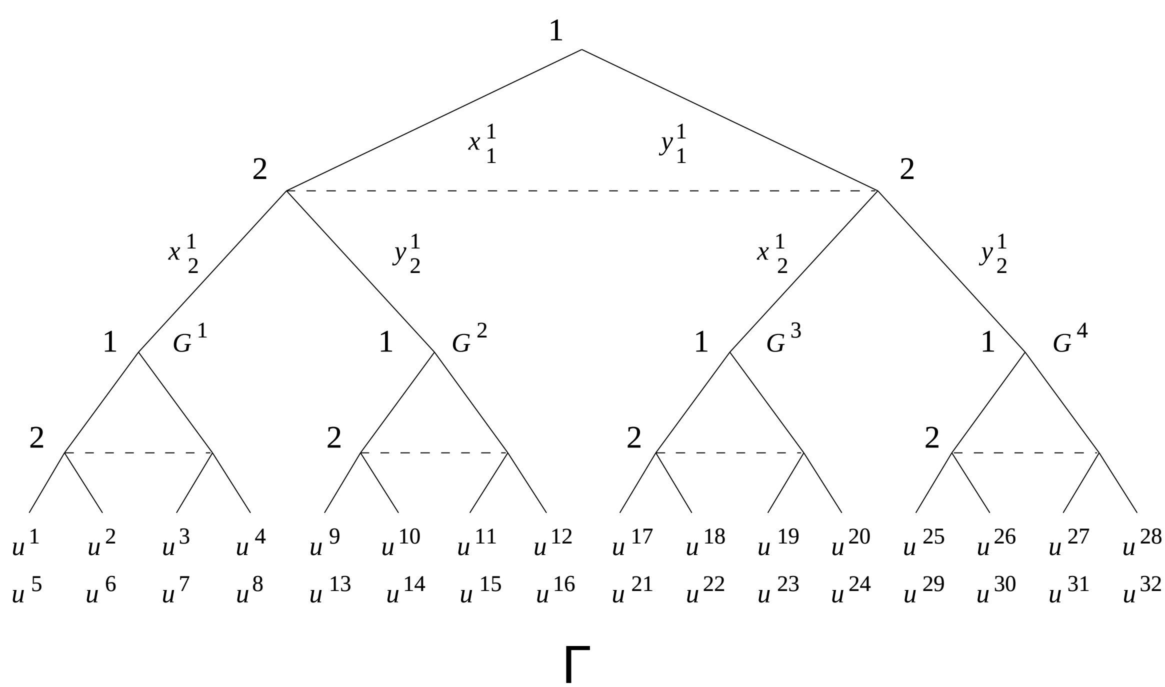

Figure 4.

Observe that the first row of u^j's are the payoffs for player 1 and the second row are payoffs for player 2. Without loss of generality, we will restrict attention to payoff vectors $u \in I\!R^{32}_{++}$. Note also that proper subgames of Γ can be viewed as one-shot games. These have been labeled G^1, G^2, G^3, and G^4 in figure 4. The set of mixed strategies for the game G^j will be denoted $\Sigma_i^j \subset I\!R^2_+$ and $\Sigma^j = \Sigma_1^j \times \Sigma_2^j$ and $\Sigma = \prod_{j=1}^n \Sigma^j$. A generic element of Σ_i^j, Σ^j, or Σ will be denoted s_i^j, s^j, or s^i respectively. Note that any payoff vector u for Γ uniquely determines a payoff vector for each game G^j. A generic payoff vector for G^j will be denoted by $u^j \in I\!R^8_{++}$, and of course $u = (u^1, u^2, u^3, u^4)$. Observe that for any given s and u, the first stage of the game $\Gamma(u)$ can also be viewed as a one-shot game where playing the action profile (x_1^1, x_2^1) yields the payoff associated with s^1 in the game $G^1(u^1)$ for the two players and (x_1^1, y_2^1) yields payoffs according to s^2 in $G^2(u^2)$, etc. We will call this game G^0. A typical payoff for G^0 will be denoted by v^0. We will use $v^0(u,s)$ to denote the implied payoff for G^0 given $u \in I\!R^{32}_{++}$ and $s \in \Sigma$. Σ_1^0, Σ_2^0, and Σ^0 are defined to be the sets of strategies for players 1 and 2 and the set of all strategy profiles for G^0, respectively.

For any $A, B \subset I\!R^m$ such that A is open and $B \subset A$, we will say that B is a generic subset of A if B is open and $\mu(A \backslash B) = 0$ where μ is the m-dimensional Lebesgue measure. Since the choice of m will always be clear from the context, we will simply write μ, instead of μ_m, μ_k, etc.

For the proof of Proposition 2, it is sufficient that $\Gamma(u)$ satisfy the following:

(i) For $j = 1, 2, 3, 4$ every best response is used with positive probability in every Nash equilibrium $s^j \in \mathrm{NE}\big(G^j(u^j)\big)$. For $j = 0$ and every $s \in \mathrm{NES}(u) = \prod_{j=1}^1 \mathrm{NE}\big(G^j(u^j)\big)$, every best response is used with positive probability in every Nash equilibrium s^0 of $G^0\big(v^0(u,s)\big)$.

(ii) For every $s = (s^1, s^2, s^3, s^4) \in \mathrm{NES}(u)$, and every $s^0 \in \mathrm{NE}\Big(G^0\big(v^0(u,s)\big)\Big)$, there exist open sets $O_1^j \subset I\!R^8_+$ and $O_2^j \subset I\!R^8_+$ and continuous functions $h_j : O_1^j \to O_2^j$ such that

$$v^0(u,s) \in O_1^0 \quad \text{and} \quad u^j \in O_1^0 \qquad \text{for } j = 1, 2, 3, 4$$

$$s^j \in O_2^j \qquad \text{for } j = 0, 1, \ldots, 4$$

$$h^0\big(v^0(u,s)\big) = s^0 \quad \text{and} \quad h_j(u^j) = s^j \qquad \text{for } j = 1, 2, \ldots, 4$$

and $h^j(x)$ is a Nash equilibrium of $G^j(x)$ for every $x \in O^j$. Moreover, the h^j's are locally unique. That is, $\bar{h}^j : \bar{O}_1^j \to O_2^j$, $\bar{O}_1^j \subset O_1^j$, $\bar{h}^j$ continuous with $\bar{h}^j(u^j) = s^j$ $\big[$or $\bar{h}^j\big(v^0(u,s)\big) = s^0\big]$ implies $\bar{h}^j$ agrees with h^j on $\bar{O}_1^j$.

Gul, Pearce, and Stacchetti [4], henceforth GPS, provide the following formulation and results.

For any normal form game G (for $v \in I\!R^{\ell n}$), let $g : I\!R^{\ell n} \times I\!R^{\ell n} \to \Sigma(G)$ be a continuous function such that $g(\hat{s}, v) = \hat{s}$ if and only if s is a Nash equilibrium of $G(v)$. GPS construct such a function g and define a strong non-degenerate Nash equilibrium $\hat{s}$ of $G(u)$ as one that satisfies the following properties:

(1) Every best response is used with positive probability in $\hat{s}$.

(2a) $\det\big(I - g_s(\hat{s}, v)\big) \neq 0$ where $g_s(\hat{s}, v)$ is the Jacobian of g with respect to s (i.e., the first $n\ell$ arguments) evaluated at $\hat{s}$.

(2b) It can be verified that if $\hat{s}$ is a strongly non-degenerate equilibrium of $G(v)$, then g is continuously differentiable in a neighborhood of $\hat{s}$ and hence there exists, by the implicit function theorem, a neighborhood O_1 of v and O_2 of $\hat{s}$ and a continuously differentiable function $h : O_1 \to O_2$ such that $h(x)$ is a Nash equilibrium of $G(x)$ for all $x \in O_1$ and $h(v) = \hat{s}$.

Moreover, this h is unique in the sense described in (ii) above. A v is called a regular payoff vector if every Nash equilibrium of $G(v)$ is strongly non-degenerate. Comparing (i) and (ii) to (1) and (2b), respectively, establishes that (i) and (ii) would be satisfied if

(I) for all $j = 1, 2, 3, 4$, u^j is a regular payoff vector for j;

(II) for all $s \in \text{NES}(u)$, $v^0(s, u)$ is a regular payoff vector for G^0.

Thus, our task is reduced to establishing the existence of a generic subset U^* of $I\!R^{32}_{++}$ such that (I) and (II) hold for every $u \in U^*$. For this, we will use two more results from GPS (result 1 is well-known). Let G be an arbitrary n-person game.

Result 1: *If v is a regular payoff vector for the game G, $G(v)$ has a finite number of Nash equilibria.*

Result 2: *There exists a generic subset V of $I\!R^{\ell n}_{++}$ such that every $v \in V$ is a regular payoff vector for G.*

Let $\bar{U}^j \subset I\!R^8_{++}$ be the generic set provided by the result 2. Obviously, $U^j := \{u \in I\!R^{32}_{++} \mid u^j \in \bar{U}^j\}$ is a generic subset of $I\!R^{32}_{++}$. Clearly, $U = \bigcap_{j=1}^4 U^j$ is also a generic subset of $I\!R^8_{++}$. For every $u \in U$, (I) is satisfied.

To establish (II), we will need the following lemma.

Lemma: *Let $f : X \to Y$ be a continuously differentiable function where $X \subset I\!R^m$, $Y \subset I\!R^k$, $k \leq m$, and X and Y are open sets. If the Jacobian f has rank k at every point $x \in X$, then for any $V \subset Y$ such that $\mu(V) = 0$, $\mu\big(f^{-1}(V)\big) = 0$.*

Proof: It is sufficient to show that $A \subset X$, $\mu(A) > 0$ implies $\mu\big(f(A)\big) > 0$. This will be done in two steps.

Step 1:

Suppose $k = m$. Then by assumption we have $\mu(A) > 0$ and $\det f'(x) \neq 0$ for all $x \in A$. Let $\mathcal{B}$ be a collection of open subsets of X such that $x \in \theta \subset X$ and θ open implies there exists $B \in \mathcal{B}$ such that $x \in B \subset \theta$. Such a collection exists; for example, let $\mathcal{B} = \{B_\epsilon(x) \bigcap X \mid x \in X,$ every coordinate of x is rational and $\epsilon > 0$ is rational$\}$ where $B_\epsilon(x)$ denotes the open ball of radius ϵ around x. To see that $x \in \theta \subset X$ and θ open implies there exists $B \in \mathcal{B}$ such that $x \in B \subset X$, note that since θ is open there exists a rational number $r > 0$ such that $B_r(x) \subset \theta$. Choose some x' with rational coordinates such that $\|x' - x\| < {}^r\!/_4$; then by the triangle inequality, $B_{r/2}(x') \subset B_r(x) \subset \theta \subset X$ and $x \in B_{r/2}(x') \bigcap X = B_{r/2}(x') \in \mathcal{B}$. By the inverse function theorem, there exists for each $x \in X$ an open set $\theta_x \subset X$ such that f is 1 to 1 on θ_x and hence there exists $B_x \in \mathcal{B}$ such that $x \in B_x \subset \theta_x$ and f is 1 to 1 on B. [Note that we are using the axiom of choice here.] Since $\mathcal{B}$ is a countable collection, so is $\mathcal{B}^* = \{B_x \mid x \in X\}$. Obviously, $\bigcup_{x \in X} B_x = X$ and since $\bigcup_{x \in X} B_x$ is the union of a countable collection of sets, it must be that $\mu\big(a \cap B_x\big) > 0$ for some B_x; otherwise, we would have $\mu(A) \leq \bigcup_{B \in B^*} \psi(A \cap B) = 0$ which is a contradiction. By the change of variable formula, we have

$$\int_{A \cap B} \big|\det f(x)\big|\, dx = \int_{f(A \cap B)} 1\, dy.$$

Since $\psi(A \cap B) > 0$ and $\big|\det f(x)\big| > 0$ for all $x \in A \cap B$, the left side of the equality is strictly positive. The right side is equal to $\mu\big(f(A \cap B)\big)$ and hence $\mu\big(f(A)\big) \geq \mu\big(f(A \cap B)\big) > 0$ as desired.

Step 2:

If $m > k$, then define $F : X \to I\!R^{m-k} \times Y$ such that $F(x^{m-k}, x^k) = \big(x^{m-k}, f(x^{m-k}, x^k)\big)$ where $x^{m-k} \in I\!R^{m-k}$ and $x^k \in I\!R^k$. Since the columns of the Jacobian of f are linearly independent, so are the columns of the Jacobian of F, i.e., $\det F(x) \neq 0$ for all $x \in X$. By step 1 above, this implies $\mu\big(F(A)\big) > 0$ whenever $\mu(A)$ is zero. Obviously, $F(A) \subset A \times f(A)$; hence, $0 < \mu\big(F(A)\big) \leq \mu(A) \times \mu\big(f(A)\big)$ and therefore $\mu\big(f(A)\big) > 0$ as desired. ▮

Let $\mathcal{B}$ be a countable collection of subsets of $I\!R^{32}_{++}$ such that for any $x \in I\!R^{32}_{++}$ and open set O such that $x \in O$, there exists $B \in \mathcal{B}$ such that $x \in O \subset \mathcal{B}$ (the existence of such a $\mathcal{B}$ is established in the proof of the lemma above).

It follows from (2b) that for any $u \in U$ and $s^j \in \mathrm{NE}(G^j)$ there exist open sets $O^j_1(s^j, u^j)$ and $O^j_2(s^j, u^j)$, and a function $h^j : O^j_1(s^j, u^j) \to O^j_2(s^j, u^j)$ such that $h^j(u^j) = s^j$, h^j is continuously differentiable and "unique" $\big($see (ii) above for a description of the sense in which h^j is unique$\big)$. By result 1, $\mathrm{NE}(G^j)$ is finite for all j. Hence, $O^j(u^j) := \bigcap_{j=1}^n \bigcap_{s^j \in \mathrm{NE}(G^j)} O^j_1(s^j, u^j)$ is an open set such that $u^j \in O^j(u^j)$. Furthermore, with each u^j we can associate a set of differentiable functions $H(u^j)$ such that each h^j corresponds to one Nash equilibrium s^j, i.e., $h^j(u^j) = s^j$. Obviously, every $h^j \in H^j(u^j)$ is continuously differentiable. Also, there exists $B(u) \in \mathcal{B}$ such that $u \in B(u) \subset \prod_{j=1}^n O^j(u^j)$. Thus, each h^j can be viewed as a continuously differentiable function on $B(u)$ to Σ^j where h^j depends only on u^j. Obviously, $\bigcup_{u \in U} B(u) = U$.

Furthermore, there are a countable number of B's in $\mathcal{B}$. Hence, U can be expressed as the countable union $\bigcup_{B \in \mathcal{B}^*} B$ for some $\mathcal{B}^* \subset \mathcal{B}$. Note also that by the "uniqueness" of h^j, $H^j(u') = H^j(u)$ for all u, $u' \in B \in \mathcal{B}^*$. Define

$$\Psi_B = \prod_{j=1}^{4} H^j(u) \qquad \text{for some } u \in B.$$

Fix any $B \in \mathcal{B}^*$ and $\psi \in \Psi_B$ and view $v^0\big(u, \psi(u)\big)$ as a function from B to $I\!R^8_{++}$. Since ψ is continuously differentiable, so is v^0. Consider the Jacobian J of v^0 evaluated at some $u \in B$. $J(u)$ will have the following form:

$$J(u) := \begin{matrix} c^1_1 & c^1_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & c^2_1 & c^2_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c^3_1 & c^3_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & c^4_1 & c^4_2 \end{matrix}$$

where each c_i^j is a column vector denoting the derivative of the payoff to player i in G^0 with respect to each of the eight components of u^j. We will show that $J(u)$ has rank 8 by showing that each pair of vectors c_1^j, c_2^j are linearly independent. Let the first four entries denote the derivatives with respect to the payoffs of player 1 and the last four entries the payoffs with respect to player 2. We claim that summing the two groups of four rows yields the 2×2 identity matrix and hence establishes the linear independence of c_1^j, c_2^j. To see this, note that increasing every payoff of a player by a constant amount leaves the Nash equilibrium unchanged and hence increases that player's own payoff by a constant amount and leaves the other player's payoff unchanged.

Thus, $J(u)$ has rank 8 for every $u \in B$. Moreover, by result 2 there exists a generic set $V^0 \in \mathbb{R}^8_{++}$ such that every $v^0 \in V^0$ is a regular payoff vector for G^0. Hence, by the lemma, ${v^0_\psi}^{-1}(V^0)$ is a generic subset of B $\big[$note that ${v^0_\psi}^{-1}(v) = u$ such that $v^0\big(u, \psi(u)\big) = v\big]$. Let $U_B = \bigcap_{\psi \in \Psi_B} {v^0_\psi}^{-1}(V^0)$; by the finiteness of Ψ_B, U_B is also a generic subset of B. Let $U^* = \bigcup_{B \in \mathcal{B}^*} U_B$. Obviously, U is open and $\mu(\mathbb{R}^{32}_{++} \backslash U^*) = \mu(\mathbb{R}^{32}_{++} \backslash U \bigcup U \backslash U^*) = \mu(U \backslash U^*) \leq \mu(\bigcup_{B \in \mathcal{B}^*} B \backslash U_B) \leq \sum_{B \in \mathcal{B}^*} \mu(B \backslash U_B) = 0$. Hence, U^* is a generic subset of $\mathbb{R}^{32}_{++}$ which establishes (II) and concludes the proof. ∎

For the case of an arbitrary stage game Γ, we would essentially have to repeat the construction of $\mathcal{B}^*$ and Ψ_B so that the payoffs associated with earlier stages can also be viewed as continuously differentiable functions of u and Nash equilibria $s(u)$ of the later stages. This becomes notationally cumbersome, but entails no novel ideas. The argument for establishing that the Jacobian of $v\big(u, s(u)\big)$ has full rank is also independent of the number of players and stages. Chance nodes would further complicate the functions $v\big(u, s(u)\big)$. In particular, even the payoffs of the final stage g "one-shot" games would now have to be viewed as weighted averages of terminal node payoffs. However, the analysis would still remain unchanged.

Proof of Proposition 2 (for the general case): The proof is by induction. Obviously, the statement is correct for $T = 1$. Suppose it is true for $T - 1$. Let Γ be a T-stage game. Let σ be the SPE in question. By the induction hypothesis, we can replace the behavior in each $(T-1)$-stage subgame of Γ with a forward induction-proof continuation such that the distribution over physical outcomes is as close to the distribution specified by σ as desired. This, together with the genericity assumption (see (ii) in the discussion of genericity above), guarantees that there exists $\bar{\sigma}$ which specifies forward induction-proof behavior in every $(T-1)$-stage subgame of Γ, is close to σ in stage 1, yields a distribution over physical outcomes that is arbitrarily close to the one associated with σ in Γ, and is an SPE of Γ^c. Again by the genericity of Γ (see (i) in the discussion of genericity), every action a_i is either used with positive probability or is threatening or irrelevant, and an action in Γ^c given $\bar{\sigma}$ is threatening iff the same action is threatening in Γ given σ.

The purpose of the remainder of this proof is to modify $\bar{\sigma}$ so that in every $(T-1)$-period subgame following an action profile $a \notin \bigcup_{i=1}^n I_i$ the behavior agrees with $\bar{\sigma}$, where I_i is the set of $(T-1)$-stage subgames of Γ that can be reached by a profile in which exactly one player i plays a threatening action and all others play according to σ, and in period 1 the behavior is close to that in $\bar{\sigma}$, and the resulting profile is forward induction-proof in the strongest sense: i.e., there are no unused relevant (and hence no threatening) strategies.

Construct γ as follows: for every $a \in I_i$, in the subgame associated with a, players play a forward induction-proof equilibrium arbitrarily close to player i's favorite SPE. This is possible by the induction hypothesis. In any subgame after $a \notin \bigcup_{i=1}^n I_i$, and in period 1, players behave according to $\bar{\sigma}$. By the definition of a threatening action, our construction of γ, and strong non-degeneracy (i.e., (i) in the discussion of genericity), every threatening action a_i yields a higher payoff for player i, against $\bar{\sigma}_{-i}$, than the strategy $\bar{\sigma}_i$. Let k denote the total number of threatening actions and let λ denote a generic element of $[0,1]^k$. Each coordinate of λ is understood to correspond to a particular threatening action. Let λ_{a_i} denote the entry of λ associated with threatening action a_i. Let $G(\bar{\sigma}, \lambda)$ denote the game obtained from $G(\bar{\sigma}), G(\gamma), \lambda$ as follows ($G(\gamma)$ for arbitrary strategy profile γ of Γ is defined in section 3): the payoffs associated with a in $G(\bar{\sigma}, \lambda)$ are the same as in $G(\bar{\sigma})$ if $a \notin \bigcup_{i=1}^n I_i$ and are equal to $\lambda_{a_i}, (1-\lambda_{a_i})$ weighting of the payoffs associated with a in $G(\bar{\sigma})$ and $G(\gamma)$, respectively, if $a_i \in I_i$. Let $G(\bar{\sigma}, \lambda, \epsilon')$ denote the *perturbation* of $G(\bar{\sigma}, \lambda)$ in which the payoff to any strategy profile s in $G(\bar{\sigma}, \lambda, \epsilon')$ is the same as the payoff associated with $\bar{s}$ in $G(\bar{\sigma}, \lambda)$ where $\bar{s}_i = (1-\epsilon')s_i + \epsilon' s'_i$ and s' is a fixed profile in which every agent's strategy places positive probability on all threatening strategies and only threatening strategies. As in the simpler case covered in section 3, if ϵ' is sufficiently small, then