Value and Perfection in Stochastic Games

Robert Samuel Simon

Faculty of Mathematics
University of Göttingen
Bunsenstr. 3-5
37073 Göttingen, Germany
simon@math.uni-goettingen.de

Center for High Performance Computing
Technical University of Dresden
Mommsenstr. 13
01062 Dresden, Germany

Abstract:

A stochastic game is *valued* if for every player k there is a function $r^k : S \to \mathbf{R}$ from the state space S to the real numbers such that for every $\epsilon > 0$ there is an ϵ equilibrium such that with probability at least $1-\epsilon$ no state s is reached where the future expected payoff for any player k differs from $r^k(s)$ by more than ϵ. We demonstrate an example of a recursive two-person non-zero-sum stochastic game with only three non-absorbing states and limit average payoffs that is not valued, (but does have ϵ equilibria for every positive ϵ). In this respect two-person non-zero-sum stochastic games are very different from their zero-sum varieties. N. Vieille proved that all such games with finitely many states have an ϵ equilibrium for every positive ϵ, and our example shows that any proof of this result must be qualitatively different from the existing proofs for zero-sum games. To show that our example is not valued we need that the existence of ϵ equilibria for all positive ϵ implies a "perfection" property. Should there exist a stochastic game without an ϵ equilibrium for some $\epsilon > 0$, this perfection property may be useful for demonstrating this fact. Furthermore our example sows some doubt concerning the existence of ϵ equilibria for two-person non-zero-sum stochastic games with countably many states.

Key words: Stochastic Games, Markov Chains, Rate of Martingale Convergence, Approximately Harmonic Functions

This research was supported financially by the German Science Foundation (Deutsche Forschungsgemeinschaft) and the Center for High Performance Computing (Technical University, Dresden).

1 Introduction

To every well defined game *value* is a natural concept. What is the maximum quantity that a player would pay to participate in that game?

An *equilibrium* is a set of strategies, one for each player, such that no player can gain in payoff by choosing a different strategy, given that all the other players do not change their strategies. A player may be willing to pay the quantity v to play a game if there is an equilibrium that gives this player at least v. A two-person game is zero-sum when the payoff to one player is always the negation of the payoff to the other player; if there is an equilibrium for a zero-sum game then there is a unique equilibrium payoff (to Player One) and it is called the *value* of the game. In general for non-zero-sum games the equilibria do not yield a unique expected payoff to each player.

For any $\epsilon \geq 0$, an ϵ-*equilibrium* in a game is a set of strategies, one for each player, such that no player can gain in payoff by more than ϵ by choosing a different strategy, given that all the other players do not change their strategies. We say that approximate equilibria exist if for every $\epsilon > 0$ there exists an ϵ-equilibrium. If a game has no equilibrium but does have approximate equilibria then there would be some advantage to some player to break from the prescribed behavior, however this advantage can be made as small as one wants. An equilibrium is often too much to expect, but approximate equilibria are the next best thing. Given finitely many players and states and a uniform bound on the payoffs, any vector cluster point of ϵ equilibrium payoffs as ϵ converges to zero will give to the game a kind of equilibrium payoff to each player. If the game is also zero-sum and has approximate equilibria then there will be a unique cluster point of ϵ equilibrium payoffs (for Player One) as ϵ converges to zero, also called the value of the game. On the other hand the lack of approximate equilibria would make irrelevant the question of how much a player is willing to pay to participate.

A *stochastic* game is played on a state space. The present state and the present behavior of all players determines stochastically the transition to a new state. All players have complete knowledge of the past history of play. A priori there is no bound on the number of stages of play. For this paper we will assume that at any given state each player has finitely many actions.

Shapley (1953) introduced the concept of a stochastic game in the context of zero-sum games where the payoffs are evaluated according to discount factors. A discount factor is a positive quantity ρ strictly less than one

such that for every stage $i \geq 0$ of play the $i+1$st stage is worth only ρ times that of the ith stage of play. The finiteness of the geometric sum $1+\rho+\rho^2+\ldots$ gives the game a compact structure where fixed point arguments are very useful. Shapley showed that such zero-sum games have equilibria and values obtainable from stationary strategies, meaning that the strategies are independent of both the stage and history of play (and are dependent only on the state).

Another way to evaluate the payoffs is by some limit of average payoffs, averaged over the stages of play. This is called the *limit average* evaluation. For example, if a player received a sequence $w_0, w_1, \ldots$ of payoffs on the stages 0, 1,..., respectively, then his payoff could be $\lim_{i\to\infty} \sup \frac{1}{i+1} \sum_{k=0}^{i} w_k$. In general, when payoffs are limit average stochastic games do not have equilibria. This was demonstrated by Blackwell and Ferguson with their famous zero-sum example "The Big Match" (1968). However they showed that this game does have approximate equilibria and a value.

Mertens and Neyman (1981) proved that every zero-sum stochastic game with limit average payoffs played on a finite state space has approximate equilibria and a value. Maitra and Sudderth (1991) extended this result to countably many states and Martin (1998) extended this result further to payoff functions defined on the infinite paths of play that are Borel with respect to their finite stage truncations.

Concerning two-player non-zero-sum games with limit average payoffs the central result was accomplished by Vieille (2000); he proved that all such stochastic game with finitely many states have approximate equilibria. For two-player non-zero-sum games with countably many states the question is still open.

If a stochastic game has finitely many players and states (and actions) and the time horizon is truncated so that there are finitely many stages then equilibria will exist (Nash 1950). The open ended nature of the time horizon gives stochastic games their most important theoretical complication. The reason that the existence of approximate equilibria does not imply the existence of an equilibrium is that the smaller the positive ϵ the longer it may take on the average for the ϵ equilibrium to be realized.

The property of "valued" for stochastic games (stated formally in the abstract) is a natural strengthening of the approximate equilibria property. Assume for now that the state space S of a stochastic game is finite. If for every positive ϵ there is an ϵ equilibrium then for every $s \in S$ and player k

there will be quantities $r^k(s)$ and a sequence of ϵ equilibria with ϵ converging to zero that will deliver average expected payoffs for player k conditioned on the first visit to s converging to $r^k(s)$ (given of course that the state s is reached with probabilities that do not converge to zero). At the second visit to any state the sub-game that remains is no different from the game that starts at this state. Of course the equilibria are only approximate, meaning that there could be a very small probability that the players find themselves in a situation calling for behavior that is far from that of an equilibrium for the future sub-game. But one could expect some property of payoff stability for some sequence of ϵ equilibria like that of the valued property, namely that if a state is reached with large probability then almost all visits to this state yield approximately the same payoffs.

In general stochastic games do not possess the valued property. With three players and finitely many states Flesch, Thuijsman, and Vrieze (1997) found a counter-example. Their game is very simple, involving only one special state where the players have any influence on the outcome of the game. They showed that for sufficiently small $\epsilon > 0$ the only ϵ equilibria involve cyclic behavior, both of the players and of the payoffs conditioned on the event that the game has remained at the special state.

We introduce a recursive two-person non-zero-sum stochastic game with three non-absorbing states and limit average payoffs that is not valued. New about our example is that the lack of the valued property is possible with only two players (and finitely many states and actions).

In Vieille's proof of approximate equilibria (for two-player non-zero-sum games with finitely many states) the behavior of the players is very complicated, involving nothing like the valued property. Naturally one wonders if there is a much simpler proof for Vieille's result, a proof closer in style to those for zero-sum stochastic games which do utilize the valued property, for example with some kind of fixed point argument on the space of payoff vectors (see Maitra and Sudderth 1991). Zero-sum stochastic games are valued; the proof is easy and provided below. Our main result shows that such an alternative approach is not possible; there is an aspect of Vieille's proof (or any alternative proof of Vieille's result) that must be complicated.

Before going further, an additional concept is necessary. A state is *absorbing* if once this state is reached the play can never leave this state, no matter what the players do. If zero is the payoff for all players at all non-absorbing states then the game is *recursive*. An absorbing state defines an isolated sub-

game, and therefore in general one associates to an absorbing state a fixed payoff for each player representing what they receive in equilibrium if this absorbing state is reached. An absorbing state in a stochastic game demonstrates vividly the difference between finitely repeated games and discounted games on the one hand and limit average games on the other. In a finitely repeated game or a discounted game when an absorbing payoff is reached the payoff to a player would be a convex combination of the absorbing state payoffs and the payoffs on the stages before this absorbing state is reached. If the payoffs to the players are limit average, once reaching an absorbing state the payoffs to the players are determined by this absorbing state only.

With the Vieille proof, the normal behavior of the players depends on more than the state which is visited. For some non-absorbing states that could be visited arbitrarily many times the players may indeed return to this state many times. Eventually a signal will be given by one of the players, and the future behavior of both players is dependent on this signal. Of course if the signal distinguishes two different ways to play such that both occur with large probability and they imply significantly different expected future payoffs to the player who is not giving the signal, then this would contradict the valued property.

Zero-sum games have a monotonicity relationship between the strategies and their values. If the payoff to a player associated with a combination of actions in a zero-sum game is increased, then the value for that player cannot go down. This is very different with non-zero-sum games and their equilibria. By increasing the payoff to some player associated with a combination of actions, an equilibrium based on cooperation may be destroyed. The introduced lack of trust and the re-establishment of balance in all new equilibria could result in lower payoffs for the player whose payoff was increased!

With our example, there is a central and initial state s and two other states t and u. At the state t Player One controls the transitions that don't lead immediately to an absorbing state (and the same holds for Player Two and the state u). For the moment we describe what happens at state t if Player Two chooses the only action that allows the game to continue with positive probability. Player One can end the game immediately, giving a good payoff to herself and and a slightly better payoff to Player Two. On the other hand she can choose a action that results in a half probability of moving back to s and a half probability of giving herself a better payoff and giving Player Two a very bad payoff. (A symmetric situation exists for

Player Two and the state u.) If Player Two trusts Player One to terminate the game with certainty from the state t then he prefers to let the game move to state t. On the other hand, with such an agreement between the players Player One discovers an advantage to choosing the move that leads back to s with half probability, given that this procedure can be repeated. If Player Two anticipates such treacherous behavior at the states t then a fight ensues at the state s concerning advancement to the state t. As a consequence of this fight Player One anticipates a bad payoff at the state s for herself, and then it would not make sense for Player One to return to s from the state t. However there are two possibilities for saving the valued property – through some form of mixed behavior by both players (as does happen with the Nash and Shapley proofs for equilibrium existence) and through punishment for choosing a forbidden action (such as the action that returns the play to s from t with half probability). Avoiding these possibilities necessitated relatively complex structures at all three states. Approximate equilibria for this game are generated by dynamic payoff solutions, for which at least some player must take turns in enjoying payoff advantage and disadvantage.

In the next section we describe the model of stochastic games with finitely many players, countably many states, and for each state finitely many actions. We establish relationships between approximate equilibria, value, and perfection. The third section contains our example and the proof that it is not valued. The fourth and last section discusses the open problem of approximate equilibria for games with countably many states.

2 The Model

We define stochastic games similarly to Markov chains and their harmonic functions. The additional complication concerns the freedom of the players to influence the transitions based on the previous history of play. By extending the definition of the state space so that distinct past histories of play leads to distinct states, we would stay close to the explicit context of Markov chains, as then a strategy choice from each player would define a Markov chain. But then we could lose track of the original structure of the game, especially if it is defined by a finite state space.

For every finite or countable set A let $\Delta(A)$ stand for the set of all probability distributions on A. If A is finite then $\Delta(A)$ is a finite dimensional

simplex. If $x \in \Delta(A)$ and $a \in A$ then the a coordinate of x will be represented as $x(a)$ (the probability given to a by x).

There is a countable or finite state space S and a finite set N of players. For every player $n \in N$ and every $s \in S$ there is a finite set A_s^n of actions. (If the action sets are countable then even for one-stage zero-sum games there are examples without approximate equilibria, for example the game where the two players choose natural numbers and the player who chooses the larger number wins a unit value.) For every $s \in S$ and every $a \in A_s := \prod_{n \in N} A_s^n$ there will be a transition law $p_a^s \in \Delta(S)$ governing the transition to states at the next stage of play after a visit to s.

We assume that the game starts at an initial state $\hat{s} \in S$. (If one prefers to start with a distribution on all the states in S one can add an initial state $\hat{s}$ that occurs only at the start of the game and such that every player has only one action at this state.) Define $\mathcal{H}_\infty := \{(\hat{s} = s_0, a_0, s_1, a_1, \ldots) \mid \forall i \geq 0 \ a_i \in A_{s_i}, \ p_{a_i}^{s_i}(s_{i+1}) > 0\}$, the set of infinite histories of play. For the initial state $\hat{s} \in S$ let $\mathcal{H}_0^{\hat{s}} := \{(\hat{s})\}$, and for every $i \geq 1$ let $\mathcal{H}_i^s$ be the set of truncations of $\mathcal{H}_\infty$ of the form $(\hat{s} = s_0, a_0, s_1, a_1, \ldots, s_{i-1}, a_{i-1}, s_i = s)$ (leaving out the actions at stage i). Let $\mathcal{H}_i$ be the union $\cup_{s \in S} \mathcal{H}_i^s$ and let $\mathcal{H}^s$ be the union $\cup_{i=0}^\infty \mathcal{H}_i^s$ (with $\mathcal{H}_0^s$ the empty set if s is not the initial state). Let $\mathcal{H}_\omega$ be the union $\cup_{i=0}^\infty \mathcal{H}_i = \cup_{s \in S} \mathcal{H}^s$. If $h \in \mathcal{H}_\omega$ is also in $\mathcal{H}^s$ then we say that h *terminates* at s.

A payoff for a player $n \in N$ in a stochastic game is a function $\mathcal{V}^n$ on $\mathcal{H}_\infty$ that is measurable with respect to the Borel σ-algebra on $\mathcal{H}_\infty$ generated by the partitions on $\mathcal{H}_\infty$ induced by the $\mathcal{H}_i$.

Usually one assumes that the payoffs for each player are uniformly bounded rather than the assumption that they are integrable in some way. The reason is simple, to define finite payoffs the integrability must be required for all strategy choices of the players. Due to the potentially great variety of strategies it is wisest in most cases to make the uniform bound assumption, which we do for the rest of this paper. Whenever applicable, let $M > 1$ be a positive real number larger than the maximal difference between all payoffs in the game.

A strategy σ^n of Player $n \in N$ is a collection of functions $(\sigma_s^n \mid s \in S)$ such that for every $s \in S$ σ_s^n is a function from $\mathcal{H}^s$ to $\Delta(A_s^n)$. For every tuple of strategies $\sigma = (\sigma^n \mid n \in N)$, one strategy for each player, probability distributions $\mu_{\sigma,i}$ are induced on the $\mathcal{H}_i$ in the natural way. We start at the initial history $(\hat{s}) \in \mathcal{H}_0^{\hat{s}}$ with $\mu_{\sigma,0}(\{(\hat{s})\}) = 1$. Given that $\mu_{\sigma,i}(h_i)$ is positive

for some $h_i \in \mathcal{H}_i^{s_i}$ and $h_{i+1} \in \mathcal{H}_{i+1}$ is a history such that the i stage truncation of h_{i+1} is equal to $h_i \in \mathcal{H}_i^{s_i}$ with $h_{i+1} = (h_i, a_i, s_{i+1})$ and $a_i = (a_i^n \mid n \in N)$ we define inductively $\mu_{\sigma,i+1}(h_{i+1}) := \mu_{\sigma,i}(h_i) p_{a_i}^{s_i}(s_{i+1}) \prod_{n \in N} \sigma_{s_i}^n(h_i)(a_i^n)$. A Borel probability distribution μ_σ is induced on $\mathcal{H}_\infty$ in the natural way, by the $\mu_{\sigma,i}$ and Kolmogorov's Extension Theorem. For every player $n \in N$ and every strategy tuple σ the distribution μ_σ generates a payoff $\mathcal{V}^n(\sigma)$ for player n as the expected value of the function $\mathcal{V}^n$ on $\mathcal{H}_\infty$, determined by the probability distribution μ_σ.

For any tuple $\sigma = (\sigma^n \mid n \in N)$ of strategies, an alternative tuple $\tilde{\sigma} = (\tilde{\sigma}^n \mid n \in N)$ and a player $k \in N$ define $\sigma|\tilde{\sigma}^k$ to be the tuple such that $\tilde{\sigma}^k$ is the strategy for player k but if $n \neq k$ then σ^n is the strategy for player n. An ϵ-equilibrium is a strategy tuple $\sigma = (\sigma^n \mid n \in N)$ such that for any alternative tuple $(\tilde{\sigma}^n \mid n \in N)$ and every player $n \in N$ it holds that $\mathcal{V}^n(\sigma|\tilde{\sigma}^n) \leq \epsilon + \mathcal{V}^n(\sigma)$.

A stochastic game is a *limit average* game when for every player $n \in N$ the Borel function $\mathcal{V}^n$ is between $\lim_{i\to\infty}\inf$ and $\lim_{i\to\infty}\sup$ of the average $\frac{1}{i}\sum_{k=0}^{i-1} w_{s_k}^n(a_k)$ where for every state $s \in S$ and $n \in N$ w_s^n is a real function defined on $A_s = \prod_{m \in N} A_s^m$.

For any tuple σ of strategies, a player $n \in N$, and a stage i of play define $v_\sigma^n : \mathcal{H}_i \to \mathbf{R}$ by $v_\sigma^n(h_i)$ equaling the expected value of $\mathcal{V}^n(\sigma)$ conditioned on reaching h_i on the ith stage, with $v_\sigma^n(h_i)$ defined to be any quantity bounded within the payoffs defining the game if the probability of reaching h_i is zero. Extend the definition of v_σ^n to $\mathcal{H}_\omega$ in the natural way. For any fixed σ and player $n \in N$ the function v_σ^n as defined on $\mathcal{H}_\omega$ with respect to μ_ω is harmonic.

Define a stochastic game to be *valued* if for every player $n \in N$ there exists a function $r^n : S \to \mathbf{R}$ such that for every $\epsilon > 0$ there is an ϵ equilibrium σ to the game such that the probability does not exceed ϵ that some history $h_i \in \mathcal{H}_i^s$ occurs with $|v_\sigma^n(h_i) - r^n(s)| > \epsilon$ for some player $n \in N$.

Proposition 1: All zero-sum stochastic games are valued.

Proof: Let $r : S \to \mathbf{R}$ be defined so that $r(s)$ is the value of the game (for Player One) starting at the state s. (The existence of this value was proven by Martin 1998). For any $0 < \epsilon \leq 1$ let σ be a strategy pair where both players guarantee a payoff (for Player One) within $\epsilon^2/10$ of the value $r(\hat{s})$, where $\hat{s}$ is the initial state. Assuming that $\max_i |v_\sigma^1(h_i) - r^n(s_i)| > \epsilon$ is obtained for a subset $\mathcal{A} \subseteq \mathcal{H}_\omega$ of finite histories reached with a probability of more than ϵ (according to the distribution μ_σ), we can assume by symmetry

that there is a subset $\mathcal{A}_1 \subseteq \mathcal{A}$ given a probability of at least $\epsilon/2$ such that $\min_i v^1_\sigma(h_i) - r^1(s_i) < \epsilon$ is obtained, (since otherwise we obtain the same for Player Two with $-r$ replacing r). We alter the strategy of Player One so that at the first stage i such that $v^1_\sigma(h_i) - r^n(s_i) < \epsilon$ Player One switches his strategy to one guaranteeing $r(s_i) - \epsilon/2$. The gain in expected payoff for Player One would be at least $\epsilon^2/4$, a contradiction. $\Box$.

For every player n define $\chi^n : S \to \mathbf{R}$ so that $\chi^n(s)$ is the min-max value for player n at the state s (the upper bound for what player n can obtain from a start at s in response to all strategy choices of the other players). The importance of the min-max value is that it represents the ability of the players to punish player n with pre-determined strategies (for example as part of an approximate equilibrium). Because the other players may be limited in their ability to coordinate their actions, this min-max value could be strictly greater than the max-min value when there are at least three players. (The max-min value is the most that player n can obtain when she must choose her strategy first and the other players respond to that strategy choice with the goal of minimizing her payoff.) Proofs that (two-person) zero-sum games have approximate equilibria demonstrate that for these games the max-min value equals the min-max value, thereafter called the value of the game.

For every $a^n \in A^n_s$ and $\hat{a} \in \prod_{k \neq n} A^k_s$ let $(\hat{a}, a^n)$ be the corresponding member of $A_s = \prod_{k \in N} A^k_s$. For every player $n \in N$ and strategy tuple σ, define the *jump* function $j^n_\sigma : \mathcal{H}^s_\omega \to \mathbf{R}$ by

$$j^n_\sigma(h) = \max_{a^n \in A^n_s} \sum_{t \in S} \chi^n(t) \sum_{\hat{a} \in \prod_{k \neq n} A^k_s} \prod_{k \in N \setminus \{n\}} \sigma^k_s(h)(\hat{a}^k) p^s_{(\hat{a}, a^n)}(t),$$

namely the maximal expected value of χ^n on the next stage following s. Extend this definition to $j^n_\sigma : \mathcal{H}_\omega \to \mathbf{R}$ in the natural way.

For any function $f : \mathcal{H}_\omega \to \mathbf{R}$, state $s \in S$, finite history $h \in \mathcal{H}^s$, action $a^n \in A^n_s$ and strategy tuple σ define $w^f_\sigma(h)(a^n)$ to be the expected value of f on the next stage after h, conditioned on the use of a^n by Player j and the use of $\sigma^k_s(h)$ by all the other players $k \neq j$. This means that

$$w^f_\sigma(h)(a^n) = \sum_{t \in S} \sum_{\hat{a} \in \prod_{k \neq n} A^k_s} f(h, (\hat{a}, a^n), t) \prod_{k \in N \setminus \{n\}} \sigma^k_s(h)(\hat{a}^k) p^s_{(\hat{a}, a^n)}(t).$$

Define $w^n_\sigma(h)(\hat{a}^n)$ to be $w^{v^n_\sigma}_\sigma(h)(a^n)$.

Definition: A strategy tuple σ of a stochastic game is *ϵ-perfect* if for every player $n \in N$ there exists a function $r^n : \mathcal{H}_\omega \to \mathbf{R}$ and a subset $\mathcal{B} \subseteq \mathcal{H}_\omega$ such that the probability of reaching $\mathcal{H}_\omega \backslash \mathcal{B}$ with the strategies σ does not exceed ϵ and for all players $n \in N$ and all finite histories $h \in \mathcal{B}$,
$r^n(h) \geq j^n_\sigma(h) - \epsilon$,
$|r^n(h) - v^n_\sigma(h)| \leq \epsilon$, and
for all actions a^n chosen with positive probability by σ^n at h
$|w^{r^n}_\sigma(h)(a^n) - r^n(h)| \leq \epsilon$.
Furthermore the strategy tuple σ is ϵ *value-perfect* if for every player n the function r^n is dependent only on the state where the history terminates. A stochastic game is *perfect* if there exists an ϵ-perfect strategy tuple for every positive ϵ. A stochastic game is *value-perfect* if there exists a ϵ value-perfect strategy tuple for every positive ϵ.

Theorem 1: A stochastic game with approximate equilibria is also perfect. A valued stochastic game is also value-perfect.

Proof: For a fixed $0 < \epsilon \leq 1$ let σ be an $\epsilon^4/(4000|N|^4M^2)$ equilibrium, and if the game is valued let $r^n : S \to \mathbf{R}$ be the value functions and let $\mathcal{A}$ be a subset of finite histories such that $\mathcal{H}_\omega \backslash \mathcal{A}$ is reached with a probability of no more than $\epsilon^4/(4000|N|^4M^2)$ and such that if h is in $\mathcal{H}_\omega \backslash \mathcal{A}$ terminating at $s \in S$ then $|v^n_\sigma(h) - r^n(s)| \leq \epsilon^4/(4000|N|^4M^2)$ for all $n \in N$. Otherwise let $\mathcal{A}$ be the empty set.

For any history $h \in \mathcal{H}^s_i$ and player n let $A^n_+(h)$ be the set of actions $a^n \in A^n_s$ chosen with positive probability according to $\sigma^n_s(h)$ such that $w^n_\sigma(h)(a^n) - v^n_\sigma(h) > \epsilon/10$ and let $A^n_-(h)$ be the set of actions $a^n \in A^n_s$ chosen with positive probability according to $\sigma^n_s(h)$ such that $w^n_\sigma(h)(a^n) - v^n_\sigma(h) < -\epsilon/10$. Let $A^n(h)$ be the union of $A^n_+(h)$ with $A^n_-(h)$. Whenever $A^n(h)$ is not empty player n could alter his strategy in the following way. If a^n is in $A^n_+(h)$ then a^n could be chosen with certainty (or any other action in this set). If a^n is in $A^n_-(h)$ then the probability given to a^n could be given to any action $\hat{a}^n$ maximizing the value of $w^n_\sigma(h)(\cdot)$. In either case we conclude that player n can increase his expected payoff conditioned on reaching h by at least $\epsilon\ \sigma^n_s(h)(a^n)/10$ (this much if a^n is in $A^n_-(h)$ and by at least $\epsilon/10$ if a^n is in $A^n_+(h)$). Because σ is an $\epsilon^4/(4000|N|^4M^2)$ equilibrium, the probability of any player ever using an action in some $A^n(h)$ according to the distribution μ_σ does not exceed $\epsilon^3/(400|N|^3M^2)$.

Let $\mathcal{D}_1$ be the union of $\mathcal{A}$ with the histories h in $\mathcal{H}_\omega$ where for some

player $n \in N$ an action in $A^n(h_i)$ is used at some history h_i resulting from the ith stage truncation of h (where $h \in \mathcal{H}_k$ and $i \leq k$). By the above we know that the probability of reaching $\mathcal{D}_1$ according to μ_σ does not exceed $\epsilon^3/(350|N|^3M^2)$.

For any history $h \in \mathcal{H}_i^s$ and player $n \in N$ let $B^n(h)$ be the set of actions $a^n \in A_s^n$ chosen with positive probability according to σ_s^n such that the probability of entering $\mathcal{D}_1$ on any following stage is at least $\epsilon/(10|N|M)$ when a^n is used against the distributions $\sigma_s^k(h)$ for $k \neq n$ and otherwise at later stages all players behave according to σ. Of course $A^n(h)$ is a subset of $B^n(h)$ (since use of an action in $A^n(h)$ causes entry into $\mathcal{D}_1$ with certainty). The probability of any player n ever using an action in some $B^n(h)$ according to μ_σ does not exceed $\epsilon^2/(35M|N|^2)$.

We define a new strategy tuple $\hat{\sigma}$. If $h \in \mathcal{H}_i^s$ and no member of $B^n(h')$ appears in h for any truncation $h' \in \mathcal{H}_k$ of h with $k < i$ then $\hat{\sigma}_s^n(h)$ is the distribution on A_s^n where the actions in $B^n(h)$ are given zero probability and the probabilities for the remaining actions are normalized, (meaning that the probability for an action $a^n \notin B^n(h)$ is $\sigma_s^n(h)(a^n)/\sum_{\overline{a}^n \notin B^n(h)} \sigma_s^n(h)(\overline{a}^n)$. If all actions are removed in this way, then define $\hat{\sigma}_s^n(h)$ to be any distribution). Otherwise, if $B^n(h)$ is empty or some member of $B^n(h')$ was played in the past then $\hat{\sigma}_s^n(h) = \sigma_s^n(h)$.

Let $\mathcal{D}_2$ be the subset of $\mathcal{H}_\omega$ such that $h \in \mathcal{D}_2$ if and only if there is some truncation h_k of h and some player n with $|v_{\hat{\sigma}}^n(h_k) - v_\sigma^n(h_k)| > \epsilon/5$. From the unlikeliness of using an action in some $B^n(h)$ we conclude that the probability of reaching $\mathcal{D}_2$ according to μ_σ does not exceed $\epsilon/7$.

Let $\mathcal{D}_3$ be the subset of $\mathcal{H}_\omega$ defined by $h \in D_3$ if and only if for some player n and some truncation $h_k \in \mathcal{H}_k$ of h the actions removed to make $\hat{\sigma}^n(h_k)$ from $\sigma^n(h_k)$ had frequency greater than $\epsilon/(5|N|M)$. Also from the unlikeliness of using an action in some $B^n(h)$ the probability of reaching $\mathcal{D}_3$ according to μ_σ does not exceed $\epsilon/7$.

Let $\mathcal{D}_4$ be the subset of $\mathcal{H}_\omega$ defined by $h \in D_4$ if and only if there is some player n and some truncation $h_k \in \mathcal{H}_k$ of h such that $v_\sigma^n(h_k) < j_\sigma^n(h_k) - \epsilon/10$. Whenever this inequality holds player n can obtain an expected payoff of at least $v_\sigma^n(h_k) + \epsilon/11$, and therefore by the approximate equilibrium property of σ the probability of reaching $\mathcal{D}_4$ according to μ_σ cannot not exceed $\epsilon^3/(300|N|^3M)$.

Define $\mathcal{B}$ to be $\mathcal{H}_\omega \backslash (\cup_{i=1}^4 \mathcal{D}_i)$, If the game is not valued let the function $r^n : \mathbf{H}_\omega \to \mathbf{R}$ for player n be v_σ^n and if the game is valued then let it be the

value function, already labeled r^n. Let $\hat{\sigma}$ be the candidate strategy tuple for the ϵ perfection property. We complete the proof with the more general v^n_σ, as the proof with the valued property is the same.

Let h be any history in $\mathcal{B}$ terminating at some $s \in S$, and let a^n be an action chosen by player n with positive probability according to $\hat{\sigma}^n_s(h)$. Due to non-membership in $\mathcal{D}_3$, when player n uses any action a^n not in $B^n(h)$ against the distributions $\hat{\sigma}^k_s(h)$ for $k \neq n$ the expected value of either χ^n or v^n_σ on the next stage does not differ by more than $\epsilon/5$ from what it would be with the distributions $\sigma^k(h)$. By non-membership in $\mathcal{D}_2$ and $\mathcal{D}_4$ this is enough for $v^n_{\hat{\sigma}}(h) \geq j^n_{\hat{\sigma}}(h) - \epsilon$. Because a^n is not in $A^n_\sigma(h)$, the expected value of v^n_σ on the next stage resulting from using a^n is within $\epsilon/10$ of $v^n_\sigma(h)$, and due to non-membership in $\mathcal{D}_2$ $v^n_\sigma(h)$ does not differ from $v^n_{\hat{\sigma}}(h)$ by more than $\epsilon/5$.

Left is to show that the probability of leaving $\mathcal{B}$ according to $\mu_{\hat{\sigma}}$ does not exceed ϵ. Because the only difference between σ and $\hat{\sigma}$ results from the use of actions in $B^n(h)$ for some $h \in \mathcal{H}_\omega$ and $n \in N$ and the probability of such an action ever being used according to μ_σ does not exceed $\epsilon^2/(35M|N|^2)$, the conclusion follows from the unlikelihood according to μ_σ of leaving $\mathcal{B}$. □

Question 1: Does there exist a stochastic game that is perfect but doesn't have approximate equilibria?

There are two problems with the converse direction of Theorem 1. First, although the probability of leaving the set $\mathcal{B}$ of histories is very small if the players stick to the suggested strategies, a player could steer intentionally the play away from $\mathcal{B}$ with unknown consequences. With regard to the proof of Theorem 1, it may be precisely the behavior of the original σ on the compliment of $\mathcal{B}$ that gave σ its approximate equilibrium property, and this could be lost with the new strategy tuple $\hat{\sigma}$. Second, given a fixed positive ϵ how do we know that there is a positive δ so small that if a player could obtain at most a payoff advantage of δ on each stage of play then this could not accumulate to an advantage of more than ϵ over time? One would like to punish a player who obtains a cumulative advantage of more than ϵ; the problem is that of testing for statistical deviation when there are arbitrarily many different situations calling for different behavior. Under what conditions the converse of Theorem 1 is possible is an interesting topic, partly because Vieille's proof (for two-person games) uses a special version of this converse direction.

Define an ϵ perfect strategy tuple σ to be ϵ *self-perfect* if the function $r^n : \mathcal{H}_\omega \to \mathbf{R}$ defining the perfection property is equal to v^n_σ. A stochastic game is *self-perfect* if there exists an ϵ self-perfect strategy tuple for every positive ϵ.

Question 2: Does there exist a stochastic game that is perfect but not self-perfect?

In the proof of Theorem 1 there may be some actions a^n of a player n at a history h in $\mathcal{B}$ terminating at s that were not removed to create $\hat{\sigma}^n_s(h)$ from $\sigma^n_s(h)$ such that $w^n_\sigma(h)(a^n)$ is very close to $v^n_\sigma(h)$, $v^n_\sigma(h)$ is very close to $v^n_{\hat{\sigma}}(h)$, but $w^n_{\hat{\sigma}}(h)(a^n)$ is far from either $v^n_{\hat{\sigma}}(h)$ or $v^n_\sigma(h)$. The action a^n may fall slightly outside the definition of $B^n(h)$ and yet there could be an action $\overline{a}^n$ by player n at a history h' following h with $v^n_{\hat{\sigma}}(h')$ far from $v^n_\sigma(h')$ such that conditioned on the use of a^n the action $\overline{a}^n$ is played later with high probability and $\overline{a}^n$ is in $B^n(h')$. In such a case the probability of player n choosing a^n at h will be small, and yet there may be no global way to choose the actions that must be removed to obtain the self-perfection property.

3 An Example

The following is an example of a recursive two-player stochastic game with limit average payoffs, three non-absorbing states and three actions for each player at each state, that is not value-perfect, meaning also that the game is not valued (by Theorem 1).

Example 1: The three non-absorbing states are the set $\{s, t, u\}$. The game is symmetric with respect to the two players, Player One and Player Two. Like a tennis match, there is a state t that means "advantage" to Player One and another state u that means "advantage" to Player Two.

At the state s both players have three actions, a for "advance", w for "wait", and c for "check". For Player $i = 1, 2$ the actions a, w, and c will be called a_i, w_i, and c_i. The transitions and payoffs are as follows.

If Player One chooses a_1 at state s and
– if Player Two chooses a_2, then the play returns to the state s,
– if Player Two chooses w_2, then the play advances to the state t,

– if Player Two chooses c_2, then the game ends with certainty with a payoff of 10 to Player One and 14 to Player Two.

If Player One chooses w_1 at state s and
– if Player Two chooses a_2, then the play advances to the state u,
– if Player Two chooses w_2, then the play returns to the state s, and
– if Player Two chooses c_2, then the game ends with a payoff of 25 to Player One and $10^{-3} = 1/1000$ to Player Two.

If Player One chooses c_1 at state s and
– if Player Two chooses a_2 then the game ends with a payoff of 14 to Player One and 10 to Player Two,
– if Player Two chooses w_2 then the game ends with a payoff of 10^{-3} to Player One and a payoff of 25 to Player Two, and
– if Player Two chooses c_2 then the game ends with a payoff of 10^{-3} to both players.

At the state t Player One has three actions, e_1 for "end", r_1 for "retreat", and f_1 for "flip". Player Two also has three actions, n_2 for "normal", b_2 for "First Bluff" and l_2 for "Second Bluff".

If Player One chooses e_1 at the state t, then no matter what Player Two does the game ends and Player One receives a payoff of 20, and
– if Player Two chooses n_2 or l_2, then Player Two receives a payoff of 21 and
– if Player Two chooses b_2 then Player Two receives a payoff of $21 + 10^{-3}$.

If Player One chooses r_1 at the state t and
– if Player Two chooses n_2, then with 1/2 probability the game ends with a payoff of 25 to Player One and 1 to Player Two and with 1/2 probability the play returns to state s,
– if Player Two chooses b_2 then the game ends with a payoff of $20 + 10^{-3}$ to Player One and 10^{-3} to Player Two, and
– if Player Two chooses l_2 then the game ends with a payoff of 20 to Player One and $7.5 + 10^{-3}$ to Player Two.

If Player One chooses f_1 at the state t and
– if Player Two chooses n_2 then with 2/7 probability the play moves to the

state u and with 5/7 probability the game ends with a payoff of 19.61 for Player One and 10^{-3} for Player Two,
– if Player Two chooses b_2 then the game ends with a payoff of 20 for Player One and $\frac{40}{7} + 10^{-3}$ for Player Two, and
– if Player Two chooses l_2 then the game ends with a payoff of $20 + 10^{-3}$ for Player One and 10^{-3} for Player Two.

At state u the situation is symmetric to that of state t, but with the roles of the players switched. Player Two has the actions e_2, r_2 and f_2 and Player One the actions n_1, b_1 and l_1.

The game starts at the state s. One should interpret the "end" of the game with corresponding payoffs to be a transition to an absorbing state. (Strictly speaking, our example has 23 absorbing states, 9 following directly after either t or u, and 5 following after s. By choosing vectors in $\mathbf{R}^2$ that contain all the payoffs in their convex hull we could reduce the number of absorbing states to three.) If the game never reaches an absorbing state then both players receive a payoff of 0.

Most critical to this example is the approximate value of 15 for a player at the state s. If Player One has an expected a payoff of 15 at s, then at the state t if Player Two chooses the action n_2 Player One will be indifferent between ending the game immediately with the action e_1 or moving back to s with 1/2 probability with the action r_1.

When analyzing this game the term *frequency* refers to the probability that an action is chosen. This is done to avoid confusion with other expressions of probability.

Lemma 1: From a start at state s each player can guarantee 10.4, meaning that $\chi^1(s)$ and $\chi^2(s)$ are at least 10.4.

Proof: By symmetry, it suffices to show that Player One can guarantee 10.4. Let Player One choose the action e_1 at state t and the action n_1 at state u. At state s let Player One choose the actions a_1, w_1, and c_1 with the frequencies .39, .26, and .35, respectively. We need to check that with the choice of any actions for Player Two that Player One receives at least 10.4 conditioned on not returning to the state s. If Player Two chooses the action c_2 then the payoff for Player One would be at least $.39 \cdot 10 + .26 \cdot 25 = 10.4$. If Player Two chooses the action w_2 then the payoff for Player One would be at least $\frac{.39 \cdot 20}{.74}$, which is more than 10.4. If Player Two chooses the actions

a_2 and r_2 then Player One would get at least $\frac{.13+.35\cdot 14}{.48}$, which is more than 10.4. If Player Two chooses the actions a_2 and f_2 then Player One would get at least $\frac{(2/7)\cdot .26\cdot 20+.35\cdot 14}{.61}$, which is more than 10.4. If Player Two chooses the actions a_2 and e_2 then Player One would get more than 16. □

It is now easy to confirm that $\chi^1(t) \geq 20$ and $\chi^2(u) \geq 20$ and that $\chi^1(u) \geq 5.7$ and $\chi^2(t) \geq 5.7$.

Theorem 2: There is no 10^{-19} value-perfect strategy pair for the game of Example 1 (implying by Theorem 1 that the game is not valued).

Before proving Theorem 2 we need some more Lemmatta, based on the assumption that there exists a 10^{-19} value-perfect strategy pair σ, with $\mathcal{B}$ the corresponding subsets of histories and $r^i : \{s, t, u\} \to \mathbf{R}$ the value functions for the players $i = 1, 2$. (For the absorbing states the value functions are already determined.) If a player i prefers some move over another or can obtain some quantity with an action we are referring either to the expected value of r^i on the next stage if that action was given a positive frequency or otherwise the expected value of χ^i on the next stage.

Lemma 2: The probability of never reaching an absorbing state from the start of the game cannot exceed 10^{-17}.

Proof: If this probability did exceed 10^{-17} then there must be a history h in $\mathcal{B}$ such that the probability of not reaching an absorbing state in the future is at least .9. Because both players can obtain a payoff of at least 5.7 from any history terminating at a non-absorbing state, the "jump" functions j^i_σ are at least 5.7 for all histories terminating at non-absorbing states, meaning also that r^i is at least $5.7 + 10^{-18}$ for all of these states and $i = 1, 2$. Since the functions r^i represent approximately what the players receive in the future from histories in $\mathcal{B}$, a member of $\mathcal{B}$ with such a low probability of future absorption would not be possible.

Lemma 3: From the start of the game the subset $\{t, u\}$ of states is reached with a probability of more than $2 \cdot 10^{-3}$.

Proof:

Case 1; $r^i(s) \geq 12.6$ for both players: It follows directly from the fact that the sums over the two players from all absorbing payoffs following directly after the state s (without first reaching either t or u) never exceed 25.001 and this sum for all absorbing payoffs of the game never exceeds

$41 + 10^{-3}$.

Case 2; $r^i(s) < 12.6$ for some player i: By symmetry we assume that $r^1(s) < 12.6$. The frequency given to a_2 at any history in $\mathcal{B}$ terminating at s does not exceed 11/12, otherwise Player One could receive at least 12.66 on this stage by playing c_1.

Let h_0 be the initial history at the first stage of play at the state s, which necessarily belongs to $\mathcal{B}$.

Case 2a; Player One chooses a_1 with positive frequency at h_0: From above we know that $\sigma_s^2(h_0)(c_2) + \sigma_s^2(h_0)(w_2) \geq 1/12$. If $\sigma_s^2(h_0)(c_2)$ were not at least twice that of $\sigma_s^2(h_0)(w_2)$ then $w_\sigma^{r^1}(h_0)(a_1)$ would be at least 10^{-18} more than $r^1(s) < 12.6$, a contradiction. Likewise $25\sigma_s^2(h_0)(w_2) \geq \sigma_s^2(h_0)(c_2)$, since otherwise $w_\sigma^{r^1}(h_0)(a_1) \geq r^1(s) - 10^{-18}$ would imply that $r^1(s)$ is less than $10.4 - 10^{-18}$, a contradiction to Lemma 1. With Player Two choosing the action c_2 with positive frequency it is necessary that $\sigma_s^1(h_0)(a_1) \geq 7/10$. But we have shown above that $\sigma_s^2(h_0)(w_2) \geq 1/312$. This implies that t is reached on the second stage with a probability of at least $\frac{1}{312}\frac{7}{10} > 1/500$.

Case 2b; Player One does not choose a_1 with positive frequency at h_0:

Clearly Player One chooses w_1 with positive frequency at h_0, since otherwise Player Two would choose w_2 with certainty. We know from Lemma 1 that $\sigma_s^2(h_0)(c_2) = 0$ (meaning that c_2 is not chosen with positive frequency). With a payoff of at least $10.4 - 10^{-18}$ for Player One from the combination of a_1 with a_2, to prevent Player One from getting at least $12.6 + 10^{-3}$ we must assume that Player Two chooses the action a_2 with a frequency of at least .6 at the history h_0. If we assume that Player One chooses the action c_1 with a frequency of at least .98 at the history h_0 then Player Two would have preferred the action w_2. Therefore we must assume that the probability of reaching u on the second stage must be at least $.6/50 = .012$. □

Lemma 4: If there are members of $\mathcal{B}$ terminating at both t and u then at any such histories the corresponding action r_i is chosen with positive frequency.

Proof: For the sake of contradiction we assume that the action r_1 is not chosen with positive frequency at some history h in $\mathcal{B}$ terminating at t. If Player Two preferred the action b_2 by more than 10^{-18} over the other two actions then only b_2 would have been chosen and then indeed Player One

would prefer the action r_1 over the others by more than 10^{-18}. Furthermore if Player Two placed all but 10^{-16} frequency on the action b_2 then the same would hold. To prevent such a preference for the action b_2 it would be necessary for $r^2(u)$ to be at least $20 + 10^{-4}$ and that Player One chooses the action f_1 with positive frequency.

Case 1; $r^1(u) \leq 20.9$**:** To prevent Player One preferring the action e_1 over the action f_1 by more than 10^{-18} at the history h it would be necessary that Player Two chooses the action l_2 with positive frequency at h. But without a positive frequency for the action r_1 this is not possible.

Case 2; $r^1(u) \geq 20.9$**:** It would be necessary that Player Two chooses the action e_2 with positive frequency for all histories in $\mathcal{B}$ terminating at u. But this implies that $r^2(u)$ is no more than $20 + 10^{-18}$, a contradiction to the above assumption. □

Lemma 5: If Player Two chooses b_2 with positive frequency at a history in $\mathcal{B}$ terminating at t it is necessary that Player One chooses r_1 with a frequency less than 1/5000. If additionally Player Two chooses both b_2 and l_2 with positive frequency then Player One chooses the action f_1 with a frequency less than 1/10000. The corresponding symmetric statement from switching the players and the states holds.

Proof: From the actions e_1 and f_1 the advantage for Player Two by choosing b_2 over the action n_2 by definition does not exceed 1/1000. Since Player Two gets at least 5.7 from the combination of r_1 with n_2, if Player One chose r_1 with a probability of at least 1/5000 Player Two would prefer by more than 10^{-18} the action n_2 over the action b_2. Additionally if Player One chose the action f_1 with a frequency exceeding 1/10000 Player Two would prefer by more than 10^{-18} the action b_2 over the action l_2. (Player Two would loose no more than $\frac{15}{10000}$ from the combination of b_2 with r_1, but then would gain at least $\frac{4}{7000} + \frac{.9997}{1000}$ from the combinations of b_2 with e_1 and f_1.) □

Lemma 6: If $r^1(s) \leq 14.9$, $r^1(u) \leq 20.9$ and there is some history in $\mathcal{B}$ terminating at t then $r^2(t)$ is at least 20.995. The corresponding symmetric statement from switching the players and the states holds.

Proof: For every history in $\mathcal{B}$ terminating at t the above assumptions imply that the combination of r_1 with n_2 gives no more than 19.95 to Player

One and the combination of f_1 with n_2 gives no more than 19.98 to Player One (with respect to the expected value of r^1 on the next stage).

First, at such a history we show that one can assume that b_2 was chosen with positive frequency. For the sake of contradiction, assume the contrary. If Player Two chose n_2 with a frequency of at least 10^{-16} then Player One would prefer e_1 over r_1 by more than 10^{-18}. Without Player One choosing r_1 with positive frequency and without Player Two choosing b_2 with positive frequency the only way to prevent Player Two from only choosing the action n_2 (which would lead quickly to a contradiction) would be that Player One chooses e_1 with a frequency of at least $1 - 10^{-18}$. But this would imply the conclusion of the lemma, and therefore we can assume that the frequency for l_2 was at least $1 - 10^{-16}$. But this leads directly to a contradiction, (as Player One would respond by choosing only f_1). Therefore we can assume that b_2 was chosen indeed with positive frequency.

Second we show that one can assume that Player Two chose l_2 with positive frequency (at any history in $\mathcal{B}$ terminating at t). Suppose the contrary. We can assume that Player One didn't choose f_1 with positive frequency, since the only way to prevent a 10^{-18} preference for e_1 over f_1 would be if all but 10^{-14} frequency went to the action b_2, (also leading to a contradiction). But with no weight given to the action f_1 the only way to prevent Player Two from choosing only the action n_2 (which would lead to a contradiction) would be for Player One to choose e_1 with a frequency of at least $1 - 1/5600$. The result of such a behavior would also imply the conclusion of the lemma.

With both b_2 and l_2 chosen with positive frequency, Lemma 5 completes the proof (for example by looking at Player Two's option to choose b_2). □

Lemma 7: From any member of $\mathcal{B}$ terminating at t the action n_2 is chosen with a frequency of at least $5 \cdot 10^{-5}$. If $r^2(s) \leq 14 + 10^{-5}$ then the action f_1 is chosen with a frequency of at least $4 \cdot 10^{-9}$. If Player One does not choose r_1 with positive frequency then the action f_1 is chosen with a frequency of at least $5 \cdot 10^{-5}$. The corresponding symmetric statement from switching the players and the states holds.

Proof: First we show that Player Two chooses the action n_2 with a frequency of at least $5 \cdot 10^{-5}$. Suppose the contrary. The first conclusion is that Player One does not choose e_1 with positive frequency, since either the action r_1 or the action f_1 would be preferable by a quantity of at least 10^{-4}. Player Two must give positive frequency to both b_2 and l_2, in the first case

to prevent Player One from choosing only the action f_1 and in the second case to prevent Player One from choosing only the action r_1 (as both would result in a contradiction). Since Player Two can get at least 20 from the state u, Player One must choose r_1 with a frequency of no more than 10^{-3}, since otherwise Player Two would prefer n_2 over b_2 by more than 10^{-18}, (as we assumed above that e_1 is not chosen with positive frequency). But then there would be no reason for Player Two to choose l_2 with positive frequency, a contradiction.

Given $r^2(s) \leq 14 + 10^{-5}$, Player Two prefers the combination of l_2 with r_1 by at least $8 \cdot 10^{-4}$ over the combination of n_2 with r_1. A frequency for e_1 above $1 - 4 \cdot 10^{-5}$ would imply that Player Two chooses only b_2, which would lead to a contradiction. Because the differences between payoffs for Player Two from the action f_1 do not exceed 8, to get Player Two to choose n_2 with positive probability (which must hold by the above) it would be necessary for Player One to choose f_1 with a frequency of at least $4 \cdot 10^{-9}$.

Lastly, if Player One did not choose r_1 with positive frequency then one must conclude that Player Two did not choose l_2 with positive frequency (since b_2 would be a much preferable action). Since Player Two can get no more than 51/7 from the combination of n_2 and f_1, it is necessary that Player One chooses f_1 with a frequency of at least $5 \cdot 10^{-4}$ to prevent Player Two from preferring b_2 over n_2 by more than 10^{-18}, (which would lead to a contradiction).

Proof of Theorem 2:

We separate the proof into three cases, determined by the use of the actions c_i.

Case I; Neither player i chooses the action c_i at any history in $\mathcal{B}$ terminating at s:

Claim: $v^i(s) \geq 20 - 10^{-7}$ for either $i = 1$ or $i = 2$ (given that Case I applies).

Because absorption is nearly certain (Lemma 2), there are histories in $\mathcal{B}$ terminating at s such that the combination of a_1 with w_2 or a_2 with w_1 are chosen with positive frequency and the corresponding state t or u, respectively, is reached with a probability of at least .499. By symmetry we assume that it was a_1 with w_2, with t reached with a probability of at least .499. If at such a history Player One chose a_1 with a frequency of less than

10^{-9} then of course $r^2(s)$ would be at least $20 - 10^{-8}$ (as Player Two could have chosen the action a_2). Likewise if Player Two chose a_2 with a frequency of less than 10^{-9} then $r^1(s)$ would be at least $20 - 10^{-8}$. Therefore from the 10^{-19} equilibrium conditions and excluding these possibilities (which imply the conclusion of our claim) we conclude from Player Two's willingness to choose the action w_2 at some history in $\mathcal{B}$ terminating at s that $r^2(s)$ is within 10^{-8} of $r^2(t)$. If there is no history in $\mathcal{B}$ terminating at s such that the action w_1 is chosen with positive frequency, then we must conclude from Lemma 2 that $r^1(s)$ is at least $20 - 10^{-8}$. Otherwise if there is such a history then we can conclude from the same logic that $r^1(s)$ is within 10^{-8} of $r^1(u)$. Due to Lemma 2 $(r^1(s), r^2(s))$ must be within 10^{-15} of a point in the convex combination of $(r^1(t), r^2(t))$ and $(r^1(u), r^2(u))$ where both $r^1(t)$ and $r^2(u)$ are at least $20 - 10^{-18}$ and $|r^2(s) - r^2(t)| \leq 10^{-8}$ and $|r^1(s) - r^1(u)| \leq 10^{-8}$. The claim is proven.

Case IA; For only one of the states t or u is there a history in $\mathcal{B}$ terminating at this state:

By symmetry we can assume that state t is reached directly from s with a probability of at least $1 - 10^{-17}$ and that $r^1(s) \geq 20 - 10^{-7}$. Due to Lemma 4 and Lemma 7 we can also assume that the probability of reaching a history terminating at t where Player One does not choose r_1 with positive probability does not exceed 10^{-8}. At any history in $\mathcal{B}$ terminating at t where Player One chooses the action r_1 with positive frequency the combination of r_1 with n_2 gives at least 20.05 for Player One and due to Lemma 7 (implying a minimal frequency for n_2) the action r_1 dominates the action e_1 by more than 10^{-18}. With Player One not choosing e_1 with positive frequency at such a history, no matter what Player Two does, we must assume that $r^2(s)$ is no more than 8, a contradiction.

Case IB; There are histories in $\mathcal{B}$ terminating at both t and u:

By symmetry we can assume that t is reached directly from s with a probability of at least .499.

Case IBi; $r^1(s) \geq 15.1$: According to Lemma 4 we know that Player One chooses r_1 with positive frequency at any history in $\mathcal{B}$ terminating at t. With the identical analysis of Case IA (concerning a comparison of r_1 with e_1) we can conclude that Player One is not choosing e_1 with positive frequency at such a history.

Now we go to an analysis of any history in $\mathcal{B}$ terminating at s such that both a_1 and w_2 are chosen with positive frequency. Since Player One does not choose the action c_1 with positive frequency nor e_1 at any history in $\mathcal{B}$ terminating at t, if Player One chooses a_1 with a frequency of at least $3 \cdot 10^{-19}$ then Player Two would not choose the action w_2 with positive frequency. So with Player One choosing a_1 with a frequency of no more than $3 \cdot 10^{-19}$ Player Two can receive a payoff of at least $20 - 3 \cdot 10^{-18}$ by choosing the action a_2 instead, and therefore $r^2(s) \geq 20 - 3 \cdot 10^{-18}$. But this is not possible, since we assume that t is reached with a probability of at least .499 and there (with the set $\mathcal{B}$) Player One does not choose e_1 with positive probability.

Case IBii; $r^1(s) \leq 15.1$: Since Player One receives at least $20 - 10^{-18}$ from the state t and this state is reached from s with a probability of at least .499, $r^1(u)$ can be no more than 10.1, at least .3 smaller than the guaranteed payoff for Player One at the state s. Furthermore there must be a history in $\mathcal{B}$ terminating at s where Player One chooses w_1 with positive frequency, since otherwise $r^1(t) \geq 20 - 10^{-18}$ would make $r^1(s) \leq 15.1$ impossible. Whenever Player One chooses w_1 with positive frequency at a history in $\mathcal{B}$ terminating at s it would be necessary that Player Two chooses w_2 with a frequency of at least $1 - 10^{-17}$. But then Player One could choose a_1 for a guaranteed payoff of at least $20 - 10^{-15}$, a contradiction to the assumption $r^1(s) \leq 15.1$.

Case II; Player i chooses the action c_i at some history in $\mathcal{B}$ terminating at s, but this does not hold for the other player:

By symmetry, we assume that it is Player One who chooses the action c_1. Since Player One can get no more than 14 with the action c_1, we have that $r^1(s) \leq 14 + 10^{-18}$. This implies for every history h in $\mathcal{B}$ terminating at s that the frequency for w_2 cannot exceed 2/5, otherwise Player One would prefer to choose the action a_1, even when punished in the event that Player Two had chosen a_2 at the same time. Likewise the frequency for a_2 must be at least $\frac{10.399}{14}$ whenever Player One is choosing c_1 with positive frequency.

Next we know that there is some history in $\mathcal{B}$ terminating at t. To contradict this claim it would be necessary that the state u is reached with a probability of at least $\frac{1}{501}$. But then Lemma 7 (its symmetric statement) would imply that t is reached from u with a probability of at least $2 \cdot 10^{-13}$, implying that indeed t is reached with a probability of at least 10^{-16}.

Case IIA; there is no history h in $\mathcal{B}$ terminating at s where Player

One chooses w_1 with positive frequency or there is no history in $\mathcal{B}$ terminating at u:

We must ask ourselves why Player Two chooses the action a_2 with a frequency of at least $3/5$ at every history in $\mathcal{B}$ terminating at s, and also look at its consequences. The probability of reaching a history where Player One chooses w_1 with a frequency over 10^{-11} cannot exceed $2 \cdot 10^{-8}$, and if Player One chooses w_1 with a frequency of less than 10^{-11} then Player One cannot choose c_1 with a frequency more than 10^{-9} (since otherwise $r^2(s) \geq 10.4$ would imply that Player Two looses too much by choosing this action). Therefore the combination of c_1 and w_2 plays no significant role in the expected payoffs and the values $(r^1(s), r^2(s))$ are within a Euclidean distance of 10^{-5} from the convex combination of $(14, 10)$ and $(r^1(t), r^2(t))$. But this is not possible, since $r^1(t) \geq 20$ and $r^1(s) \leq 14 + 10^{-18}$, which would imply that $r^2(s)$ is less than 10.1, (a contradiction to Lemma 1).

Case IIB; there is some history in $\mathcal{B}$ terminating at s where Player One chooses the action w_1 with positive frequency and there is some history in $\mathcal{B}$ terminating at u:

Since Player Two chooses the action a_2 with a frequency of at least $3/5$ at any history h in $\mathcal{B}$ terminating at s, we must assume that $|r^1(u) - r^1(s)| \leq 10^{-17}$. But then by Lemma 6 we have $r^2(t) \geq 20.995$, implying that there was no reason for Player Two to choose the action e_2 at any history in $\mathcal{B}$ terminating at u (since by Lemma 4 and Lemma 7 at all such histories Player Two chooses f_2 and r_2 with positive frequency). This implies that both $r^1(u)$ and $r^1(s)$ are below 8, a contradiction to Lemma 1.

Case III; Both Players $i = 1, 2$ choose the action c_i at some history in $\mathcal{B}$ terminating at s:

We start with $r^i(s) \leq 14 + 10^{-18}$ for both $i = 1, 2$. Lemmatta 3 and 7 imply that there are histories in $\mathcal{B}$ terminating at t and at u. Also from $r^i(s) \leq 14 + 10^{-18}$ for both $i = 1, 2$ we conclude that both $r^1(t)$ and $r^2(u)$ do not exceed 20.1.

Furthermore, Player One must chose a_1 with a frequency of at least $1/50$ at any history in $\mathcal{B}$ terminating at s, and the same must hold for Player Two and a_2. The proof for this is simple. If Player One chose c_1 with a frequency of more than .28 and also a_1 with less than $1/50$ then Player Two could get at least 14.1 from responding with the action w_2 (even if he is punished for choosing this action). If Player One chose w_1 with a frequency of more than

.7 then Player Two could get at least $14 + 10^{-3}$ by choosing the action a_2.

Also by Lemma 3 there is some history h in $\mathcal{B}$ terminating at s where the combination w_i and a_k are played, with $i \neq k$. Without loss of generality we can assume that the actions a_1 and w_2 were played together at some history h in $\mathcal{B}$ terminating at s.

Case IIIA; the action w_1 was not chosen with positive probability at the history h: With a frequency of at least 1/50 given to a_1, if $r^2(t)$ were at least $14 + 10^{-16}$ then $w^{r^2}(h)(w_2)$ would be at least $14 + 2 \cdot 10^{-18}$, which would contradict the main assumptions of Case III. From Lemma 6 we must conclude that $r^1(u)$ is at least 20.995.Therefore from Lemma 7 (establishing a minimal frequency for the move n_2) Player One prefers the action f_1 over the action e_1 by more than 10^{-18} at any history in $\mathcal{B}$ terminating at t, implying that $r^2(t)$ is less than 8. Due to the minimal frequency of 1/50 given to a_1 (and Player Two's positive frequency for w_2) we must presume (at the history h) that Player One also chose c_1 with a frequency of at least 1/1000. Since Player One is not choosing w_1 with positive frequency Player Two could not have chosen the action a_2 with positive frequency, a contradiction.

Case IIIB; the action w_1 was chosen with positive probability at the history h:

Since both a_1 and a_2 were chosen with a frequency of at least 1/50 we must assume (by the same argument of Case IIIA) that $r^1(u)$ and $r^2(t)$ are both no more than $14 + 10^{-16}$. On the other hand, by Lemma 6 we must assume that both $r^1(u)$ and $r^2(t)$ are at least 20.995, a contradiction. q.e.d.

We know of at least two ϵ equilibrium strategies to this game.

For the first approximate equilibrium, at all visits to s let the players alternate between playing a_1 with w_2 and playing a_2 with w_1. When at the state t let Player Two choose only the action n_2 and let Player One choose e_1 with a frequency of 9/21 and r_1 with a frequency of 12/21. Let the players act symmetrically at the state u. The future expected payoffs for their visits to the state s will alternate between $(15, 20)$ and $(20, 15)$, which will imply that Player One will be indifferent between his two actions at the state t and the same holds for Player Two at the state u. It is easy to check that neither player has any motivation to deviate.

The second equilibrium corresponds to the Vieille proof, and it is not so easy to find. Most of the time both players perform the action a at the state

s. At the state t Player Two will choose n_2 with certainty and Player One will choose the actions r_1 and e_1 with frequencies 7/13 and 6/13, respectively. At the state u Player One will choose the action n_1 with certainty and Player Two will choose the actions r_2 and e_2 with frequencies 6/13 and 7/13, respectively. Every time the play is at the state s the players count the number of times the play has been consecutively in s. Let N be any natural number greater than $100/\epsilon$. If it is the Nth consecutive visit to s or something less then Player Two plays a_2 with certainty. In those initial N stages Player One chooses at each stage the action w_1 with a positive frequency p which satisfies $(1-p)^N = 5/6$ (and otherwise a_1 is chosen). If the players reach the $N+1$st consecutive visit to s then Player One will play a_1 with certainty and Player Two will choose w_2 with the frequency of a very small quantity δ (for example $\delta < \epsilon/40$ suffices) and c_2 also with a frequency of δ. They continue in this way until either w_2 or c_2 is chosen. One can calculate that the expected payoffs for both players at a first visit to s will be 15, Player One gets always an expected payoff of 15 at states s and u, and Player Two receives a expected payoff of 14 from the combination of w_2 and a_1. The expected payoff for Player Two at the state s varies within each extended visit to that state, starting at 15 and falling to 14 if Player One fails to choose w_1. With δ sufficiently small Player One cannot gain more than ϵ by choosing w_1 when Player Two is choosing w_2 or c_2 with the very small frequency δ.

4 Conclusion: Countably Many States

The discovery of the above example originated from our curiosity concerning the existence of approximate equilibria for two-player stochastic games with countably many states. How complicated must be the equilibria for two-person stochastic games with finitely many states could be very relevant for two-person stochastic games with countably many states. The main reason for our pessimism concerning the existence of approximate equilibrium for the latter class of games lies precisely in the very complicated way approximate equilibria are realized in Vieille's proof for the former class of games. The rate of convergence of the average payoffs played a minimal role in Vieille's proof. If for some strategy pair the number of expected visits to any non-absorbing state is finite then almost surely an absorbing state will be reached, and this

fact was used critically in the Vieille proof. The pigeon hole principle lies behind this claim, using the finiteness of the number of non-absorbing states. If there is a proof of the existence of approximate equilibria for two-person games with countably many states, one could expect that its application to the finite state space games would deliver ϵ equilibria with strong properties concerning the rate of convergence of the average payoffs.

Notice with our example that the first equilibrium involved a fast process of reaching an absorbing payoff and the second equilibrium (that following the Vieille proof) involved a slow process of reaching an absorbing state, (although the rate of convergence could be made much faster if the strategies were altered to incorporate punishment of a player who chooses an action that was given zero frequency). We suspect that the first equilibrium is ad-hoc, meaning that it does not belong to a class of approximate equilibria whose existence is guaranteed.

If an analogy to a different area of mathematics is allowed (not completely irrelevant due to the use of the pigeon hole principle in Vieille's proof), in Ramsey Theory there are some contexts in which the existence of infinite monochromatic structures are implied and other contexts where they are not implied. If one colors the 2 subsets of a countable set with two colors then there will be an infinite subset with monochromatically colored pairs. On the other hand, although a two-coloring of the integers yields arbitrarily long monochromatic arithmetic progressions, one can two-color the integers so that there is no infinite monochromatic arithmetic progression. The reasoning is simple: With a list $l_1, l_2, \ldots$ of all the (countably many) infinite arithmetic progressions, choose a sequence $(a_1, a_2, \ldots)$ of numbers such that a_i appears in l_i for every $i \geq 1$ and such that $|a_i| \geq 3^i$ for every $i \geq 1$. This is possible because the l_i are infinite progressions. Color the sequence $(a_1, a_2, \ldots)$ red and color the rest of the integers blue.

There is an additional equilibrium concept more problematic when there are infinitely many states, that of uniformity. An ϵ equilibrium for a limit average stochastic game is *uniform* if for some sufficiently large N for all $n \leq N$ it defines ϵ equilibria for the truncated games that end at the nth stage (and where the payoffs are determined by the averages over the n stages). Vieille's proof is that of the existence of uniform ϵ equilibria for every positive ϵ. We choose to consider a class of games for which the existence of the approximate equilibria implies that there are uniform ϵ equilibria for every $\epsilon > 0$. Assume a stochastic game is recursive, such that every absorbing

state has fixed payoffs for all players (for example at every absorbing state every player has only one action) and some player can obtain a positive $w > 0$ from a start at any non-absorbing state. With any ϵ equilibrium the probability that an absorbing state is not reached from the start of the game cannot exceed ϵ/w, and therefore there must be a stage K such that the probability of absorption before reaching the stage K is greater than $1 - 2\epsilon/w$. Letting N be greater than $\frac{K}{\epsilon}$, the original ϵ equilibrium defines $4\epsilon(M+1)/w$ equilibria for all games of n stages with $n \geq N$ (where M was the bound on the maximal difference between any two payoffs).

A strategy for finding a counter-example could be following. Construct an infinite sequence of recursive games $\Gamma_0, \Gamma_1, \ldots$ corresponding to increasing finite sets $S_0 \subseteq S_1 \subseteq \ldots$ of non-absorbing states such that for every $i \geq 0$ and $j \geq i$ the actions are the same for all games Γ_j, all payoffs for both players from absorbing states are greater than one, and if a is an action tuple at $s \in S_i$ then $p(i)^s_a(t) = p(j)^s_a(t)$ for all $j \geq i$ and $t \in S_i$, (with $p(k)$ the transition laws for the game Γ_k). Construct a game played on a countable state space ($\cup_{i=0}^{\infty} S_i$ unioned with absorbing states) by having the game start at $\hat{s} \in S_0$, define the non-absorbing states on the ith stage to be the set S_i, and declare that absorption occurs on stage i if an absorbing state of the game Γ_i is reached. Furthermore, give both players the ability to force the game to the set of absorbing states with certainty for a start at any non-absorbing state of the new countable state space. Desirable may be games Γ_i such that with large i the approximate equilibrium behavior of Γ_i keeps the non-absorbing play most of the time close to the set S_0 and the minimal number of stages necessary to reach an absorbing state in the game Γ_i starting from any $s_0 \in S_0$ goes to infinity as i goes to infinity. Otherwise if we allow that absorbing states are reachable quickly from all non-absorbing states it may be desirable to redefine the game so that reaching an absorbing state of Γ_i on the ith stage of play does not mean certain absorption but rather a positive probability of absorption mixed with a positive probability of starting the game over again at $\hat{s} \in S_0$.

5 References

BLACKWELL, D. and FERGUSON, T.S. (1968). "The Big Match", *Annals of Mathematical Science*, **39**, pp. 159-163.

FLESCH, J., THUIJSMAN, F., and VRIEZE, O.J. (1997). "Cyclic Markov Equilibria in a Cubic Game", *International Journal of Game Theory*, **26**, pp. 303-314.

MAITRA, A. and SUDDERTH, W. (1991). "An Operator Solution of Stochastic Games", *Israel Journal of Mathematics*, **78**, pp. 33-49.

MARTIN, D. (1998). "The Determinacy of Blackwell Games", *The Journal of Symbolic Logic,* **63**, pp. 1565-1581.

MERTENS, J.F. and NEYMAN, A. (1981). "Stochastic Games", *International Journal of Game Theory*, **10**, 53-66.

NASH, J. (1950). "Equilibrium Points in n-Person Games", Proceedings of the National Academy of Sciences, **36**, pp. 48-49.

SHAPLEY, L.S. (1953). "Stochastic Games", *Proc. Nat. Acad. Sci. U.S.A.*, **39**, pp. 1095-1100.

VIEILLE, N. (2000). "Two-Player Stochastic Games I: A Reduction", *Israel Journal of Mathematics*, **119**, pp. 55-91, "Two-Player Stochastic Games II: The Case of Recursive Games", *Israel Journal of Mathematics*, **119**, pp. 92-126, "Small Perturbations and Stochastic Games", *Israel Journal of Mathematics*, **119**, pp. 127-142.