Discounted Repeated Games Having Computable Strategies with No Computable Best Response under Subgame-Perfect Equilibria

Dargaj, Jakub; Simonsen, Jakob Grue

Published in:
ACM Transactions on Economics and Computation

DOI:
10.1145/3505585

Publication date:
2022

Document version
Peer reviewed version

Document license:
Unspecified

Citation for published version (APA):
Discounted Repeated Games having Computable Strategies with no Computable Best Response under Subgame-Perfect Equilibria

JAKUB DARGAJ and JAKOB GRUE SIMONSEN, Department of Computer Science, University of Copenhagen, Denmark

A classic result in computational game theory states that there are infinitely repeated games where one player has a computable strategy that has a best response, but no computable best response. For games with discounted payoff, the result is known to hold for a specific class of games – essentially generalizations of Prisoner’s Dilemma – but until now, no necessary and sufficient condition is known. To be of any value, the computable strategy having no computable best response must be part of a subgame-perfect equilibrium, as otherwise a rational, self-interested player would not play the strategy.

We give the first necessary and sufficient conditions for a two-player repeated game $G$ to have such a computable strategy with no computable best response for all discount factors above some threshold. The conditions involve existence of a Nash equilibrium of the repeated game whose discounted payoffs satisfy certain conditions involving the minmax payoffs of the underlying stage game. We show that it is decidable in polynomial time in the size of the payoff matrix of $G$ whether it satisfies these conditions.

CCS Concepts: • Theory of computation → Computability; • Applied computing → Economics.

Additional Key Words and Phrases: repeated games; discounted payoff; best response strategies; Nash equilibria; subgame-perfect equilibria

ACM Reference Format:

1 INTRODUCTION

We consider pure strategies in two-player games $G$ with perfect information. In a repeated game (or supergame), $G$ is played repeatedly with each player $P_i \in \{P_1, P_2\}$ aware of all moves played by all players in all previous repetitions. In discounted such games, in each successive repetition of the game, the payoff for each player is discounted by some fixed factor $\delta$ with $0 < \delta < 1$, and the payoff after an infinite number of repetitions is the sum over all discounted repetitions. A computable strategy for infinitely repeated games is one where an algorithm computes the next action based on the finite history of previous repetitions of the game. It has been known for more than 25 years that particular infinitely repeated games admit computable strategies that have a best
response—a sequence of actions played by the opposing player that maximizes the opponent’s payoff against the strategy—but no computable best response [17, 22]. In algorithmic terms: some algorithm will play a strategy such that there will exist a counterstrategy for the other player that will achieve maximal payoff among all strategies, but no such counterstrategy is computable. Note that non-computability of a strategy is an absolute property of the strategy and is not contingent on it being a response to some other strategy. To be credible, a strategy played by some player has to be one that a rational player aiming to maximize their payoff would actually play, hence be part of a subgame-perfect equilibrium—the appropriate notion of Nash equilibrium for infinitely repeated games.

While the above results are known for particular games, typically Prisoner’s Dilemma or some generalizations of it, we provide necessary and sufficient conditions for any two-player game with perfect information to have the property that there exists a computable strategy with no computable best response (even though a best response strategy exists) with discounted payoff. A consequence of this result is that it is decidable—indeed decidable in polynomial time—whether a game has the property. Moreover, it is easy for a game to have it, and many standard examples of games do.

For infinitely repeated games with limit-of-means payoff, we have very recently obtained complete characterizations of games with computable strategies with no computable best response under Nash and subgame-perfect equilibria using the key insights that (i) players’ minmax payoff in punishment phases in trigger strategies can mimic cooperation and defection from Prisoner’s Dilemma in almost all games, and (ii) that to establish an equilibrium, strategies from the Folk theorems (standard results describing the set of equilibrium payoffs) can be carefully modified by incorporating recursively inseparable sets in select repetitions of the game [10]. However, for infinitely repeated games with discounted payoff, existence of computable strategies with no computable best response have been proven for games that, essentially, are generalizations of Prisoner’s Dilemma [22], but no necessary and sufficient conditions for games to have such strategies exist in the literature. While the key insights of [10] are applicable to the discounted setting, the constructions are necessarily more complicated compared to limit-of-means payoff. This is because a computable strategy $s$ that attempts to diagonalize against all other computable strategies must simulate all other machines, entailing that deviations by an opponent from a prescribed path of play cannot in general be detected immediately (indeed, can usually only be detected when a Turing machine implementing the opponent’s strategy has been simulated on sufficiently many inputs). Thus, the strategy $s$ will in general have a delay in detecting deviations, and unlike limit-of-means payoffs—discounted payoffs are exponentially decreasing with the length of the delay.

The aim of the present paper is to give a complete characterization of the infinitely repeated two-player games where one player has a computable strategy without computable best response under discounted payoff, and show that this characterization leads to a polynomial-time algorithm to decide whether a game has this property.

1.1 Contributions

We give the first necessary and sufficient conditions for any infinitely repeated two-player game $G$ with discounted payoffs to have a computable strategy with no computable best response. As with several existing theorems in repeated games with discounted payoffs$^2$, we prove that the

---

$^1$For discounted games, the set of pure strategies is compact in the product topology, and as the function mapping strategy profiles to payoffs is continuous, it hence obtains a maximum, see e.g. [22].

$^2$Typical examples are the Folk Theorem for discounted games, see e.g. [13],[20], or Theorem 2.19 in this paper, and the previous results for computable strategies with no computable best response [22].
result holds for all discount factors $\delta$ larger than some threshold $\delta$. Specifically, for any two-player normal-form game $G$ we find necessary and sufficient conditions on $G$ such that its infinitely repeated version $G^\infty$ satisfies that there exists some $\delta$ with $0 < \delta < 1$ such that for any $\delta$ with $\delta < \delta < 1$, there is a (pure) strategy-profile $s = (s_1, s_2)$ such that the following three conditions hold:

1. $s$ is a subgame-perfect equilibrium of $G^\infty$ with $\delta$-discounted payoff.
2. $s_2$ is a computable strategy.
3. For any best response $s'_1$ to $s_2$ in $G^\infty$ with $\delta$-discounted payoff, $s'_1$ is not a computable strategy.

Our main result (Theorem 3.24) yields necessary and sufficient conditions for $G$ to satisfy the above. Roughly, the conditions involve existence of a feasible and strictly individually rational payoff profile—a convex combination of the stage game payoffs where each player obtains strictly more than their minmax payoff—but with a snag: There are games $G$ where, for every discount factor $\delta$, every payoff of one player in a Nash equilibrium in $G^\infty$ equals $1/(1 - \delta)$ times their minmax payoff in $G$—we call such games shallow—and in some shallow games, every computable strategy has a computable best response (see Example 3.19); we show that if a (computably) reduced version—the canonical reduction—of $G$ satisfies the minmax criterion, we recover the main result. The resulting characterization is reminiscent of the conditions in Folk Theorems for repeated games (see, e.g., Theorem 2.19).

Formally, for any $G$, let $G_\delta^\infty$ denote its infinitely repeated version with $\delta$-discounted payoff. Our main result is:

**Theorem (Theorem 3.24).** Let $G = (\{1, 2\}, A, u)$ be a 2-player normal-form game, and let $G^{\text{min}} = (\{1, 2\}, A^{\text{min}}, u^{\text{min}})$ be its canonical reduction. The following conditions are equivalent:

(a) There is a $\delta$ with $0 < \delta < 1$ such that for all $\delta$ with $\delta < \delta < 1$ there is a strategy profile $s = (s_1, s_2)$ such that the following hold:

1. $s$ is a subgame-perfect equilibrium of $G_\delta^\infty$.
2. $s_2$ is computable,
3. $s_2$ does not have a computable best response.

(b) There is a feasible and individually rational payoff profile $\nu'$ in $G$ that satisfies:

$$u'_1 > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2),$$

and at least one of the following two conditions:

$$u'_2 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_2(a_1, a_2),$$

$$u'_1 > \min_{a_2 \in A_2^{\text{min}}} \max_{a_1 \in A_1^{\text{min}}} u_1^{\text{min}}(a_1, a_2).$$

In addition to the above, we show (Section 3.7) that there is no computable search procedure for computing a best response when given a computable strategy as input.

Our results subsume the prior results for discounted games [22] and give the first complete characterization of all 2-player games having computable strategies without computable best responses under discounted payoffs. The results can be seen as strong confirmation of the implicit conjecture made by Nachbar and Zame [22] who wrote that the assumption that games satisfy

---

3 The sufficient conditions of [22] imply that condition (b) of Theorem 3.24 holds, and Examples 3.25 and 3.27 contain games that do not satisfy the assumptions of [22], but do satisfy the conditions in (b) (and hence also (a)) of Theorem 3.24.
their sufficient conditions “greatly facilitates our argument, but is probably not necessary for our conclusions”.

To illustrate the applicability of Theorem 3.24, consider the well-known games Prisoner’s Dilemma and Rock-Paper-Scissors in Examples 1.1 and 1.2 below.

**Example 1.1 (Prisoner’s Dilemma).** Let \( a, b, c, d \in \mathbb{R} \) satisfy \( c > a > d > b \). Prisoner’s Dilemma is a two-player game with \( A_1 = A_2 = \{C, D\} \) and the following payoff matrix:

\[
\begin{array}{cc}
C & D \\
C & a, a \quad b, c \\
D & c, b \quad d, d
\end{array}
\]

The minmax payoff profile of Prisoner’s Dilemma is \((d, d)\), and the payoff profile \((a, a) > (d, d)\) by the definition, thus (b) of Theorem 3.24 holds, and each player has a computable strategy with no computable best response (as already known from [22]).

**Example 1.2 (Rock-paper-scissors).** Rock-paper-scissors is a two-player game with \( A_1 = A_2 = \{\text{Rock, Paper, Scissors}\} \) and payoff matrix as follows:

\[
\begin{array}{ccc}
\text{Rock} & \text{Paper} & \text{Scissors} \\
\hline
\text{Rock} & 0, 0 & -1, 1 \\
\text{Paper} & 1, -1 & 0, 0 \\
\text{Scissors} & -1, 1 & 0, 0
\end{array}
\]

The minmax payoff profile of Rock-paper-scissors is \((1, 1)\). However, as 1 is also the maximum payoff each player can earn, no combination of payoff profiles gives any player payoff greater than 1, so the condition (b) of Theorem 3.24 is not satisfied.

We give several further examples of games \( G \) that satisfy some, but not all, of the conditions of (b) in Section 3.6.

We also prove that it is decidable in polynomial time in the size of the payoff matrix of \( G \) whether \( G \) has a computable strategy without computable best response. Indeed, we have:

**Theorem (Theorem 4.1).** It is decidable in polynomial time in \(|A_1| \times |A_2|\) whether a two-player normal-form game \( G = \{\{1, 2\}, A = A_1 \times A_2, u\} \) satisfies conditions (b) of Theorem 3.24.

We stress that like previous work in the area, our results concern pure strategies and exact best responses: strategies are deterministic and non-probabilistic, and a best response has to be obtained, not just approximated. It is not hard to see that for any \( \epsilon > 0 \) and any computable strategy \( s_2 \) for P2, there is a computable strategy for P1 such that (the payoff of) playing \( s_1 \) against \( s_2 \) \( \epsilon \)-approximates the best response to \( s_2 \), and thus neither existing results [22], nor ours, carry over to the approximate setting.

### 1.2 Main ideas and overview of the proof

The top-level approach is a standard technique in repeated games: P2 plays a strategy \( s_2 \) that defines a prescribed path of play, that is, an infinite sequence of actions to be played in the repetitions \( G \) that both players should follow to maximize their discounted payoffs; by slowly diagonalizing against all computable strategies, \( s_2 \) will force the prescribed path to induce a subgame-perfect equilibrium for some non-computable strategy for P1, but such that no computable strategy \( s_1 \) is a best response to \( s_2 \) in the repeated game. Any deviation from the prescribed path of play results in infinite punishment (a grim trigger), forcing both players into a low-payoff subgame-perfect equilibrium.
The construction of the prescribed path and two punishment paths follows the same general route as Nachbar and Zame’s similar results for generalizations of the Prisoner’s Dilemma [22], and we use the same mechanics (diagonalization via Smullyan’s notion of recursively inseparable sets) to ensure the absence of computable best responses. However, the games of Nachbar and Zame necessarily satisfy stronger conditions than ours\(^4\), and our construction must delicately wrangle all possibilities for Nash equilibria where the payoffs of either player might dominate, or equal the minmax payoffs of the stage game \(G\). In the absence of actions with specific properties available, we construct finite sequences of actions (that are guaranteed to exist for a larger class of games) simulating cooperation and defection in the Prisoner’s Dilemma, and employ a Folk Theorem construction to obtain payoff profiles for reward and punishment sequences. Moreover, as Folk Theorems for discounted games are incomplete for payoffs equal to the minmax payoffs we are forced to analyze such games in detail (Section 3.3).

The high-level idea is to define a strategy that will split a path of play of \(G^\infty\) into repetitions of finite sequences of rounds of \(G\), with each repetition having length \(K = T + K_r + K_a\). Here, \(T\) is the length of a test period needed to ensure that P1 plays a particular finite sequence of actions, \(K_r\) is the length of a reward period where P2 will play actions that ensure that both players obtain high discounted payoff for playing according to the prescribed path, and \(K_a\) is the length of an adjustment period used to decrease the maximum difference in payoffs until a potential deviation is detected, so that punishment is still effective. Proving that P1 would receive less payoff by straying from the prescribed path is done by a classical application of the One-Shot Deviation Principle (Lemma 3.14).

The three positive integers \(T\), \(K_r\), and \(K_a\) are, in effect, parameters of the strategy \(s_2\), and setting these parameters judiciously is a major part of the proof. The length, \(T\) of the test period, and the prescribed sequence of actions in it, is chosen according to a finite history of play that is ensured to exist by the assumptions on the existence of certain Nash equilibria (Lemma 3.3); and the parameters \(K_r\) and \(K_a\) are chosen by deriving a finite set of lower bounds for \(K_r\) and \(\delta\) above which the main result is ensured to hold, and setting \(K_a\) accordingly (Section 3.2). Together, this ensures existence of a computable strategy \(s_2\) that is part of a subgame-perfect equilibrium, but has no computable best response. The converse implication, namely that existence of such a strategy implies existence of Nash equilibria of \(G^\infty\) with the desired properties follows much more readily (Lemma 3.1) using well-known techniques.

For polynomial-time decidability of the conditions in our main result, we employ a Folk Theorem characterization of the set of Nash equilibria combined with straightforward geometric arguments applied to both the original stage game and its canonically reduced version.

Remark 1 (A note on real and rational numbers). As usual in most of the literature on computability in repeated games (e.g., as in[7, 16, 17, 19, 22]) we use real numbers in definitions and statement of results, with the understanding that the underlying Turing machines use sufficiently good rational approximations in all places. As we are concerned with proving statements that assert existence of real numbers such that various inequalities hold, it is easy to see that all such inequalities hold when reals are replaced by \(\epsilon\)-close rationals (but that best responses remain exact, rather than just approximate). It is likely that all of our results—and indeed that many of the results in the literature on computability in repeated games—can be done for Turing machines working directly with representations of real numbers, for example Type-2 Turing machines [31] where arbitrary—not necessarily computable—real numbers can be fed as input streams to an otherwise ordinary Turing machine; however, this is beyond the scope of the present paper.

\(^4\)Indeed, [22] presupposes actions \(C_1, D_1 \in A_1\) and \(C_2, D_2 \in A_2\) such that \((D_1, D_2)\) is a Nash equilibrium of \(G\), but the payoff from \((C_1, C_2)\) strictly Pareto dominates \((D_1, D_2)\).
1.3 Related work

The original impetus for studying computable strategies with no best response was studied in the setting of limit-of-means payoff for infinitely repeated games. In that setting, no discounting occurs, but instead a player’s payoff after infinitely many rounds is the liminf of the average payoff after finitely many repetitions; Knoblauch proved that the Prisoner’s Dilemma admitted computable strategies that have a best response, but no computable best response [17] under limit-of-means payoff, a result later improved by Fortnow and Whang [12] showing that there is a polynomial-time computable strategy in Prisoner’s Dilemma that has no eventually $\epsilon$-optimal computable response for any $\epsilon > 0$. For discounted payoffs, Nachbar and Zame showed that, there are computable strategies with best responses where no best response is computable for a class of two-player games that are paradoxical in the same way as the Prisoner’s dilemma—rational players earn less in one round than if they were both forced to make an irrational decision [22]. Unlike previous results for limit-of-means payoff, the strategies in [22] are required to be subgame-perfect equilibria, and the conditions for existence of strategies without computable best responses are sufficient, but the authors conjecture that they are not necessary.

Both prior to, and after, the landmark results of Knoblauch and Nachbar and Zame, substantial work has been devoted to computing best responses (or Nash equilibria) for repeated games where strategies are constrained to be computable by machines with less power than the full Turing machines. Classic work includes Rubinstein [27], Gilboa [15], Ben-Porath [3], and Neyman and Okada [23] (finite automata); Fortnow and Whang [12] (polynomial-time computable strategies). Modern results have mostly concerned variations on the notion of equilibria or asymmetry between players, for example Chen et al. consider strategies with strictly bounded memories (a setting slightly different from strategies computable by finite automata) [8], and Zuo and Tang [32] study Stackelberg equilibria in a setting with restricted machines, and Chen et al. [9] study changes to Nash equilibria of infinitely repeated games under restrictions on the running time or space of the Turing machines. For games with discounted payoff, [4] prove that all subgame-perfect equilibrium paths consist of elementary subpaths that can be represented as directed graphs.

Similar results concerning notions different from strategies that are known to exist classically, but fail to be computable, exist elsewhere in Economics; for example, Richter and Wong showed that there are exchange economies with all components computable and where a competitive equilibrium exists (by the Arrow-Debreu Theorem [1]), but no such equilibrium is computable [25].

Furthermore, substantial attention has been given to the study of the complexity of finding Nash or subgame-perfect equilibria in repeated games with and without discounting. Classic constructions (known as "Folk Theorems") typically characterize the set of possible payoff profiles for Nash and subgame-perfect equilibria in a way that suggests that equilibria should be easy to compute; indeed, for two-player games with limit-of-means payoff, Littman and Stone showed that a Nash equilibrium can be found in polynomial time [19]. However, for three or more players, Borgs et al. showed that finding both exact and $\epsilon$-approximate Nash equilibria are PPAD-hard for certain discount factors, but as proven by Halpern, Pass, and Seeman [16], changing the underlying model of computation slightly (considering polytime Turing machines maintaining state across previous computations and assuming common cryptographics constraints), finding $\epsilon$-approximate Nash equilibria can be done in polynomial time.

Finally, Dargaj and Simonsen gave a complete characterization of the set of two-player games that have a computable strategy without computable best response in the setting of limit-of-means payoff [10]. In the discounted case, the transition from the general Prisoner’s Dilemma to any normal-form game is not as straightforward as for the limit-of-means case [10] because (i) the Folk Theorem for discounted games is weaker, and (ii) the characterization includes a lower bound on
the discount factor. However, the conditions in the characterizations of both the discounted and the limit-of-means case are quite similar, and involve the ability of each player to earn (strictly) more than their minmax payoff.

1.4 Some open problems

We prove our results for two-player games with pure strategies and perfect information. Any of these requirements could be relaxed, but it is unclear how difficult it would be to obtain similar complete (and decidable) characterizations of games with strategies having no computable best response. For games with 3 or more players, there is the possibility that our proof methods – already delicate in meting out punishment for one opponent – might suffer from the same issues that make the Folk Theorem for multi-player games more complete—e.g., for subgame-perfect equilibria, the strongest version of the Folk Theorem for discounted games can only ensure characterization of exact rather than approximate equilibria via the set of feasible and individually rational payoff profiles if the set is full-dimensional [13, 20], and it is possible that a similar requirement could prevent a sufficient and necessary characterization in the style of the present paper.

On a different note, our results show that the existence of computable strategies without computable best responses is ensured for all discount factors above some threshold \( \delta \). Just as (proofs of the) Folk Theorem, our techniques rely on the discount factor being large enough that short initial sequences of repeated play do not contain so much payoff that a player cannot credibly threaten with grim-trigger-style punishment. However, the exact value of \( \delta \) is of interest to see when the phenomenon appears; however, even for the Folk Theorem—the proof of which shares technique with the ones we use, but is substantially less delicate in its construction—finding such threshold values even for symmetric games with very small sets of action profiles requires fairly involved analysis [5].

Finally, there is the question of how well computable response strategies for P2 can approximate the uncomputable best response in case P1 plays a strategy with no computable best response. For games with limit-of-means payoff, Fortnow and Whang showed that for Prisoner’s Dilemma, there are computable strategies such that, for any \( \epsilon \) with \( 0 < \epsilon < 1 \), no computable response is eventually \( \epsilon \)-optimal [12]. However, this result is patently false under discounted payoffs; indeed, for any computable strategy \( s_2 \) for P2 and any \( \epsilon > 0 \), there is a computable strategy \( s_1 \) for P1 that \( \epsilon \)-approximates the best response [22], and this strategy is uniformly computable in \( s_2 \) and \( \epsilon \) in the sense that there is a program \( P \) that, given \( \epsilon \) and program computing \( s_2 \), \( P \) outputs \( s_1 \) simply by optimizing over simulations of \( s_2 \) over sufficiently long finite horizons (until the “tail” of further repetitions can only yield less than \( \epsilon \) payoff in total). However, for strategies with bounded computational resources, e.g. strategies constrained to run in polynomial time, we expect the situation to be more complex: we conjecture that while for every polynomial-time computable strategy \( s_2 \) and every \( \epsilon > 0 \) there is a computable \( \epsilon \)-approximation \( s_1 \) to a best response to \( s_2 \), such a strategy cannot be uniformly obtained, that is, there is no program that on input \( s_2 \) and \( \epsilon \) will output a computable \( \epsilon \)-approximation.

2 PRELIMINARIES

We expect the reader to be familiar with basic notions from game theory and computability theory at the level of introductory textbooks (e.g., [14, 18, 24, 28]). To keep the paper self-contained, we recapitulate notation and some fundamental results in the following. Even though we are primarily interested in two-player games, we give definitions for games with any finite number of players in order to conform to standard notation. We set \( \mathbb{N} = \{1, 2, \ldots\} \), \( \mathbb{N}_0 = \{0\} \cup \mathbb{N} \), and we denote the sets of rational and real numbers by \( \mathbb{Q} \) and \( \mathbb{R} \), respectively. If \( m \) and \( n \) are integers, we write \( m | n \) if \( m \) divides \( n \). If \( X \) is a finite sequence and \( Y \) is a finite or infinite sequence, we write \( X \preceq Y \) if \( X \) is a
prefix of $Y$. We write the concatenation of $X$ and $Y$ as $XY$, and if $t_1, t_2$ are both positive integers with $t_1 \leq t_2 < |X|$, we write $X[t_1]$ for the element at index $t_1$ in $X$ and $X[t_1 : t_2]$ for the contiguous subsequence of $X$ starting at $t_1$ and ending at $t_2$.

2.1 Game theory

Definition 2.1 (Normal-form game). A (normal-form) game is a tuple $(N, A, u)$ where:

1. $N = \{1, \ldots, n\}$ is the set of players (referred to as $P_1, \ldots, P_n$).
2. $A = A_1 \times \cdots \times A_n$ is the set of action profiles, where $A_i$ is a finite set of actions available to $P_i$.
3. $u = (u_1, \ldots, u_n)$, where $u_i : A \rightarrow \mathbb{R}$ is the payoff (aka. utility or reward) function for $P_i$.

A payoff profile is any element of $\mathbb{R}^n$.

In this paper, we are specifically interested in the case $N = \{1, 2\}$.

Definition 2.2 (Pareto domination). Let $G = (N, A, u)$ be a normal-form game. Action profile $a$ is said to Pareto dominate action profile $a'$ if (i) For all $i \in N$, $u_i(a) \geq u_i(a')$, and (ii) there is $i \in N$ such that $u_i(a) > u_i(a')$. If, for all $i \in N$, we have $u_i(a) > u_i(a')$, we say that a strictly Pareto dominates $a'$.

For example, in Prisoner’s Dilemma the action profile $(C, C)$ strictly Pareto dominates $(D, D)$. This follows from the initial assumption that for both players $i$, $u_i(C, C) = a > d = u_i(D, D)$.

Definition 2.3. For an action profile $a = (a_1, \ldots, a_n)$ and $P_i$, we denote by $a_{-i}$ the tuple of actions of all other players, that is $(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$.

Definition 2.4 (Best response in stage games; (pure) Nash equilibrium). Let $G = (N, A, u)$ be a game, let $a = (a_1, \ldots, a_n) \in A$ be an action profile, and let $a'_i \in A_i$ be an action of $P_i$. We say that $a'_i$ is a best response to $a_{-i}$ if $u_i(a'_i, a_{-i}) \geq u_i(a'_i', a_{-i})$ for any other action $a'_i' \in A_i$. We say that $a$ is a (pure) Nash equilibrium of $G$ if, for all $i \in N$, $a_i$ is a best response to $a_{-i}$.

Example 2.5 (Stag hunt). Stag hunt is a two-player game $G$ with $A_1 = A_2 = \{\text{Stag, Hare}\}$ and the following payoff matrix:

\[
\begin{array}{c|cc}
 & \text{Stag} & \text{Hare} \\
\hline
\text{Stag} & 3, 3 & 0, 2 \\
\text{Hare} & 2, 0 & 1, 1 \\
\end{array}
\]

Stag hunt is an example of a game that does not satisfy the sufficient conditions stated in [22] – the only Nash equilibrium of $G$ is $(\text{Stag, Stag})$ with $u(\text{Stag, Stag}) = (3, 3)$ and clearly, no action profile in $G$ strictly Pareto dominates $(\text{Stag, Stag})$. However, the minmax payoff profile of $G$ is $(1, 1)$ (as well as of the canonical reduction $G^{\min} = G$), and the payoff profile $(3, 3) > (1, 1)$ is feasible. Hence, all three conditions of (b) of Theorem 3.24 are satisfied and consequently, $P_2$ has a computable strategy without a computable best response.

2.2 Repeated games

We consider the situation when the same game is played infinitely many times; standard treatments of such games can be found in [2, 14, 21], and we recapitulate basic terminology here.

Definition 2.6 (Infinitely repeated game). Given a game $G = (N, A, u)$, $G^\infty$ is a game which consists of infinitely many repetitions of the game $G$. $G$ is called the stage game of the infinitely repeated game $G^\infty$.

Next we define a finite history of length $T \in \mathbb{N}_0$ as a sequence of the first $T$ action profiles played in $G^\infty$ and a path of play as infinite sequence of action profiles.
Similarly, we define the normalized $\delta$-discounted payoff, $G^\infty_\delta$. Given an infinite sequence of payoffs $u = \bar{u}_1[1], \bar{u}_1[2], \ldots$ for Pi and a discount factor $\delta$ with $0 < \delta < 1$, the $\delta$-discounted payoff of Pi is:

$$v_i (u, \delta) = \sum_{t=1}^{\infty} \delta^{t-1} \bar{u}_1[t].$$

Similarly, we define the normalized discounted payoff of Pi to be

$$\bar{v}_i (u) = (1 - \delta)v_i (u, \delta)$$

For a stage game $G$ and a real number $0 < \delta < 1$, we denote by $G^\infty_\delta$ the infinite game $G^\infty$ where the payoff of any infinite path of play is given by the $\delta$-discounted payoff.

For simplicity of notation, we shall usually suppress $\delta$ in $G^\infty_\delta$ when used in statements that are independent of $\delta$, for examples in most statements on histories, strategies, or subgames.

When the value of $\delta$ is small, the first few stages affect the discounted payoff significantly. For example, in the infinitely repeated Prisoner’s Dilemma with $\delta < 1/2$, the first stage contributes to the infinite sum possibly with $c$, while the maximum any player can obtain in the rest of the game is $\frac{\delta}{1-\delta}c < c$.

Note that as the payoff in each instance of the stage game is bounded and $0 < \delta < 1$, the sequence of partial sums $\left(\sum_{t=1}^{T} \delta^{t-1} \bar{u}_1[t]\right)_{T=1}^{\infty}$ is convergent, and thus $v_i (u, \delta)$ is well-defined.

Using the normalized discounted payoff allows to directly compare the payoffs of the stage game $G$ and its discounted infinite version $G^\infty_\delta$, it is easy to see that obtaining a payoff of 1 in each repetition of the stage game obtains a (normalized) discounted payoff in $G^\infty_\delta$ of 1.

The action played by a player in the stage $t+1$ depends on the history of length $t$. All players have complete information about the actions played before, so a player’s strategy maps finite histories into actions played in the next stage:

**Definition 2.10 (Strategy in a repeated game).** Let $G = (N, A, u)$ be a game. A (pure) strategy for Pi in $G^\infty$ is a map $s_i : \mathcal{H}_{G^\infty} \rightarrow A_i$. A strategy profile in $G^\infty$ is a tuple $s = (s_1, \ldots, s_n)$ where, for each $i \in N$, $s_i$ is a strategy for Pi.

A strategy $s_i$ is said to be computable if there is a Turing machine that, on input any finite history $h \in \mathcal{H}_{G^\infty}$ halts with output $s_i(h)$.
Observe that any strategy profile \( s \) defines a unique path of play \( h^\infty \), namely the one where each player in stage \( t \in \mathbb{N} \) of \( G^\infty \) observes the finite history consisting of actions played by all players in stages \( 1, \ldots, t - 1 \), and then uses their strategy to play an action for stage \( t \). If \( s = (s_1, \ldots, s_n) \) is a strategy profile and \( s'_i \) is a strategy for \( Pi \), we write \((s'_i, s_{-1})\) for the strategy profile obtained by replacing \( s_i \) by \( s'_i \).

**Definition 2.11 (Payoff of a strategy profile).** Let \( G = (N, A, u) \) be a game, let \( s = (s_1, \ldots, s_n) \) be a strategy profile in \( G^\infty \), and let \( h^s_i \) be the unique path of play induced by \( s \). The discounted payoff of \( Pi \) is:

\[
u_i(s) = \sum_{t=1}^{\infty} \delta^{t-1} u_i(h^s_i[t])
\]

It is natural to view strategies \( s \) in \( G^\infty \) as actions in a stage game, and the definitions of all standard notions carry over mutatis mutandis. For notational convenience, we give the full definition of best response and Nash equilibrium below.

**Definition 2.12 (Best response; Nash equilibrium in \( G^\delta \)).** Let \( G = (N, A, u) \) be a game, let \( s = (s_1, \ldots, s_n) \) be a strategy profile in \( G^\delta \), and let \( s^*_i \) be a strategy for \( Pi \) in \( G^\delta \). We say that \( s^*_i \) is a best response to \( s_{-i} \) if \( v_i(s'_i, s_{-i}) \geq v_i(s^*_i, s_{-i}) \) for any other strategy \( s'_i \) for \( Pi \). We say that \( s \) is a Nash equilibrium if, for all \( i \in N, s_i \) is a best response to \( s_{-i} \).

For a two-player game and a strategy profile \( s = (s_1, s_2) \) we abuse notation slightly by writing that \( s_1 \) is a best response to \( s_2 \) instead of a best response to \( s_{-1} = (s_2) \). Observe that no player can unilaterally choose an action (or strategy) that yields them a strictly better payoff than a Nash equilibrium—any strictly better payoff must involve other players changing strategies as well.

It is a standard result that every strategy in a discounted game has a best response (see, e.g., [22] for a discussion):

**Theorem 2.13.** Let \( G \) be a 2-player stage game and \( 0 < \delta < 1 \) a real number. Then, every strategy in \( G^\infty \) has a best response.

### 2.3 Subgames and subgame-perfect equilibria

**Definition 2.14 (Subgame).** Let \( G^\infty \) be an infinitely repeated game, \( T \in \mathbb{N} \) and \( h^T \in \mathcal{H}^T_{G^\infty} \). The subgame \((G^\infty, h^T)\) is the infinitely repeated game starting at stage \( T + 1 \) of \( G^\infty \) with history \( h^T \).

To illustrate the notion of a subgame, consider a \( G^\infty \) and a strategy profile \( s \) inducing the path of play \( h^s_\infty \). If the history \( h^T \) is a restriction of \( h^s_\infty \) to the first \( T \) stages, then \( s \) applied to the subgame \((G^\infty, h^T)\) leads to the path of play \( h^\infty [T+1, \ldots] \), where \( h^\infty [T+1, \ldots] \) is the contiguous subsequence of \( h^s_\infty \) starting at stage \( T + 1 \). On the other hand, there may be histories containing actions that, according to \( s \), are never played by any of the players. Every such history defines a different subgame, and leads to a path of play that may have nothing in common with the original \( h^s_\infty \).

As a subgame \((G^\infty, h^T)\) together with a strategy profile \( s \) leads to a unique path of play, we can define the payoff for each player in the subgame \((G^\infty, h^T)\) induced by \( s \). This is called the continuation payoff:

**Definition 2.15 (Continuation payoff).** Let \( T \in \mathbb{N}_0 \) and \( h^T \in \mathcal{H}^T_{G^\infty} \). Assume that a strategy profile \( s \) in the subgame \((G^\delta, h^T)\) leads to the path of play \( h^\infty \in \mathcal{H}^\infty_{G^\infty} \). The continuation payoff (aka. the present value of the payoff in the stage \( T + 1 \)) of \( Pi \) is:

\[
u_i(s; h^T) = \sum_{t=1}^{\infty} \delta^{t-1} u_i(h^\infty[t])
\]
The conclusions of all folk theorems are approximately the same: every payoff profile satisfying $p_\text{\textit{\alpha}}$ Maskin-Fudenberg argument in Appendix A with computability explicit.

Involving three or more players [7, 11, 16].

see [11]; in general, both the statement and the (computational) usefulness of Folk Theorems are more complex for games involving three or more players [7, 11, 16].

Let us now focus on strategy profiles with strategies ignoring the histories and playing a Nash equilibrium every time. In the infinitely repeated Prisoner’s Dilemma, this could be $s = (s_1, s_2)$ with $s_1(h^T) = s_2(h^T) = D$ for every $h^T \in H_{G^{\infty}}$. Since the future actions of $P_i$ are independent of the actions in stage $T + 1$ and of the subgame, $P_i$’s best response in the stage $T + 1$ is to maximize their payoff in the current stage. This is summed up in the following observation.

**Observation 1.** Let $a \in A$ be a Nash equilibrium of a stage game $G$ and $0 < \delta < 1$ a real number. Then the repeated play of $a$ is a subgame-perfect equilibrium of the infinitely repeated game $G_{\delta}^{\infty}$.

The following theorem is an adaption of its general version first stated in [6].

**Theorem 2.17 (One-shot deviation principle).** Strategy profile $s = (s_1, \ldots, s_n)$ is a subgame-perfect equilibrium of the infinitely repeated game $G_{\delta}^{\infty}$ if and only if for each $P_i$, each stage $T$ and history $h^T \in H_{G^{\infty}T}$, no deviation from $s_i(h^T)$ would raise $P_i$’s payoff.

### 2.4 The Folk Theorem

Folk theorems characterize the payoff profiles that are achievable under equilibria in different settings, depending on how the payoff is computed or which kind of equilibria we are interested in. The conclusions of all folk theorems are approximately the same: every payoff profile satisfying two minimal requirements is achievable under an equilibrium. First, individual rationality, demands at least the obvious minimal payoff for every player and second, feasibility, ensures that the payoffs in the repeated game can be combined from the stage game payoffs. The proofs of folk theorems are usually constructive, and thus provide us with actual strategy profiles that lead to given payoff profiles.

**Definition 2.18.** Let $G = ([1, 2], A, u)$ be a normal-form game and let $v = (v_1, v_2) \in \mathbb{R}^2$ be a payoff profile.

1. $v$ is individually rational if, for each $i \in \{1, 2\}$: $v_i \geq \min_{a_i \in A_i} \max_{a_i \in A_i} u_i(a_i, a_{\text{\textit{-i}}})$.
2. $v$ is strictly individually rational if, for each $i \in \{1, 2\}$: $v_i > \min_{a_i \in A_i} \max_{a_i \in A_i} u_i(a_i, a_{\text{\textit{-i}}})$.
3. $v$ is feasible if there exists a vector $\alpha \in \mathbb{R}_{\text{\textit{\geq}}}^{|A|}$ with non-negative components satisfying $\sum_{a \in A} \alpha_a = 1$ and $\forall i \in \{1, 2\}$: $v_i = \sum_{a \in A} \alpha_a u_i(a)$.

Observe that if $p, q \in \mathbb{R}$ are non-negative, satisfy $p + q = 1$, and if $v$ and $v$ are both feasible, then $pv + qv$ is feasible.

We present the Folk Theorem for 2-player games. The original proof [13] allows mixed strategies and needs a public randomization device. In a later paper, Maskin and Fudenberg [20], using a fully constructive proof, showed that both of these conditions are unnecessary, and we present their result for pure strategies below.

**Theorem 2.19 (The Folk Theorem).** Let $G^{\infty}$ be an infinitely repeated 2-player game and $v$ be a normalized payoff profile.

5 Note that the Folk Theorem for discounted games is more problematic when there are more than two players involved, see [11]; in general, both the statement and the (computational) usefulness of Folk Theorems are more complex for games involving three or more players [7, 11, 16].

6 For clarity, we have made computability explicit in the statement of the theorem, and provide a proof sketch of the Maskin-Fudenberg argument in Appendix A with computability explicit.
We now continue to the proof of the main result. First, we have the following lemma.

(1) For every $\delta$ with $0 < \delta < 1$, if $v = \frac{1}{(1-\delta)^2}$ is a payoff profile under a Nash equilibrium of $G_\delta^{\omega}$, then $v$ is feasible and individually rational.

(2) For any $\epsilon > 0$ there is $\delta < 1$ such that for any $v$ that is feasible and strictly individually rational and for any discount factor $\delta$ with $\delta < \delta' < 1$ there is a computable subgame-perfect equilibrium $s$ of $G_\delta^{\omega}$ with the payoff profile $v(s) = \frac{1}{(1-\delta')^2}$. Moreover, the normalized continuation payoffs after every history $h^T$ satisfy $\bar{v}(s; h^T) < \bar{v} + \epsilon$.

### 2.5 Computability theory

As usual, for any $A \subseteq \mathbb{N}$ we say that $A$ is recursively enumerable if there is a Turing machine that halts exactly on the elements of $A$ (equivalently, outputs exactly the elements of $A$), and that $A$ is decidable if there exists a Turing machine that halts on all inputs and accepts on input $n$ if $n \in A$.

**Definition 2.20.** We assume a standard Gödel numbering of the Turing machines and denote by $T_m$ the $m$th Turing machine in this numbering, and by $\phi_m : \mathbb{N} \rightarrow \mathbb{N}$ the partial function computed by $T_m$. If $n \in \mathbb{N}$, we write $\phi_m(n)\downarrow$ if $T_m$ halts on input $n \in \mathbb{N}$. The jump is the set $0' = \{n \in \mathbb{N} : \phi_n(n)\downarrow\}$.

The jump $0'$ is known to be recursively enumerable and undecidable [26, §13.1]. We shall use Smullyan’s notion of recursive inseparability [29]:

**Definition 2.21.** Let $\Sigma$ be a non-empty alphabet. Sets $A, B \subseteq \Sigma^*$ are said to be recursively inseparable if $A \cap B = \emptyset$ and there is no decidable set $C \subseteq \Sigma^*$ such that $A \subseteq C$ and $B \subseteq \Sigma^* \setminus C$.

Observe that if $A$ is not decidable then $A$ and its complement are recursively inseparable. We use two standard sets known to be recursively inseparable:

**Definition 2.22.** Define $\mathcal{A} = \{n \in \mathbb{N} : \phi_n(n)\downarrow \land \phi_n(n) = 0\}$, and $\mathcal{B} = \{n \in \mathbb{N} : \phi_n(n)\downarrow \land \phi_n(n) \neq 0\}$.

The following is well-known and provable by standard methods (see, e.g. [22]):

**Proposition 2.23.** Sets $\mathcal{A}, \mathcal{B}$ and $\mathcal{A} \cup \mathcal{B}$ are recursively enumerable and undecidable. Furthermore, $\mathcal{A}$ and $\mathcal{B}$ are recursively inseparable.

**Definition 2.24.** For $n \in \mathbb{N}$, define:

$$
\mathcal{A}_n = \{i \in \mathbb{N} : (i \leq n) \land (T_i \text{ halts in at most } n - i \text{ steps on input } i) \land (\phi_i(i) = 0)\}
$$

$$
\mathcal{B}_n = \{i \in \mathbb{N} : (i \leq n) \land (T_i \text{ halts in at most } n - i \text{ steps on input } i) \land (\phi_i(i) \neq 0)\}
$$

**Remark 2.** Observe that $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subseteq \mathcal{A}_3 \subseteq \cdots \subseteq \mathcal{A}$ and $\mathcal{B}_1 \subseteq \mathcal{B}_2 \subseteq \mathcal{B}_3 \subseteq \cdots \subseteq \mathcal{B}$. Clearly, $\mathcal{A}_n$ and $\mathcal{B}_n$ are finite for all $n \in \mathbb{N}$ and hence decidable (even stronger: there exists a Turing machine that on input $n$ will output (the Gödel number of) a Turing machine deciding $\mathcal{A}_n$ because a universal Turing machine can simulate up to $n$ steps of $T_i$ on input $i$; similarly for $\mathcal{B}_n$). Observe also that for $n \in \mathcal{A}$, there is some $k \in \mathbb{N}$ such that $T_n$ halts in $k$ steps on input $n$, whence $n \in \mathcal{A}_{n+k}$.

### 3 PROOF OF THE MAIN RESULT

We now continue to the proof of the main result. First, we have the following lemma.

**Lemma 3.1.** Let $G$ be a 2-player normal-form game and let $0 < \delta < 1$. If there exists a strategy profile $s = (s_1, s_2)$ in $G_\delta^{\omega}$ satisfying

1. $s$ is a Nash equilibrium of $G_\delta^{\omega}$,
2. $s_2$ is a computable strategy,
3. $s_2$ does not have a computable best response,
then there is a Nash equilibrium \( s' \) of \( G_{\delta}^\infty \) satisfying:

\[
v_1(s') > \frac{1}{1 - \delta} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2). \tag{1}
\]

**Proof.** Suppose that every Nash equilibrium \( s' \) leads to a payoff profile \( v(s') \) such that

\[
v_1(s') \leq \frac{1}{1 - \delta} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2).
\]

We prove that any strategy profile that satisfies conditions (1) and (2) cannot satisfy condition (3). Let \( s = (s_1, s_2) \) be a Nash equilibrium of \( G_{\delta}^\infty \) where \( s_2 \) is a computable strategy. Define \( \tilde{s}_1 \) to be the strategy of P1 that, given a finite history \( h^T \in H_{G_\infty}^T \), in stage \( T + 1 \) computes \( \tilde{a}_2 = s_2(h^T) \) and plays an action \( \tilde{a}_1 \) such that \( u_1(\tilde{a}_1, \tilde{a}_2) = \max_{a_1 \in A_1} u_2(\tilde{a}_1, \tilde{a}_2) \geq \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) \). Now,

\[
v_1(\tilde{s}_1, s_2) \geq \sum_{t=1}^{\infty} \delta^{t-1} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) = \frac{1}{1 - \delta} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) \geq v_1(s_1, s_2),
\]

and as \( s_1 \) is a best response to \( s_2 \), \( \tilde{s}_1 \) is also a best response to \( s_2 \). Moreover, \( s_2 \) was assumed to be computable, so \( \tilde{s}_1 \) is clearly computable, and thus condition (3) is violated. \( \square \)

Rock-paper-scissors from Example 1.2 has no Nash equilibrium and the same holds for its infinitely repeated version, so the requirements of Lemma 3.1 are trivially not satisfied. The following simple example illustrates a game that has a Nash equilibrium, but no player can reach more than their minmax payoff, and hence the inequality (1) in Lemma 3.1 is violated for any Nash equilibrium.

**Example 3.2.** Let \( G \) be the two-player game with \( A_1 = A_2 = \{x, y\} \) and payoff matrix below.

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>1, 1</td>
<td>0, 0</td>
</tr>
<tr>
<td>y</td>
<td>0, 0</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

Even though the set of Nash equilibria of \( G_{\delta}^\infty \) is infinitely large for any \( \delta \), they all have the normalized payoff profile \( (1, 1) = \left( \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2), \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2) \right) \).

Next, Lemma 3.3 and Theorem 3.4 concern games that have a Nash equilibrium with strictly individually rational payoff profile for both players. Later in the paper we show how to modify the arguments used in the proofs in the case where there is no Nash equilibrium with strictly individually rational payoff profile for P2 (i.e., P2 cannot earn more than their minmax under a Nash equilibrium).

The statement of Lemma 3.3 may appear fairly technical, but the intuition is that the lemma establishes existence of particular finite histories and strategies such that (1) the finite histories mimic single stages in Prisoner’s dilemma, and (2) the strategies have properties analogous to actions in Prisoner’s dilemma, enabling Theorem 3.4 to be proved by techniques similar to the proof of [22] for the special case Prisoner’s Dilemma.

**Lemma 3.3.** Let \( G \) be a 2-player normal-form game, and let \( 0 < \delta' < 1 \). Assume that there is a Nash equilibrium \( s \) of \( G_{\delta'}^\infty \) that satisfies both (a) and (b) below

(a) \[ v_1(s) > \frac{1}{1 - \delta'} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) \]
Observe that in Figure 1:

\[(\mathcal{G}^0, \mathcal{H}^0, \mathcal{T}^0) \in \mathcal{H}^0 \text{ such that the following hold:}
\]

\[(1)\quad v(h^T_C) = \sum_{t=1}^{T} \delta^{t-1} u(h^T_C[t]) > \sum_{t=1}^{T} \delta^{t-1} u(h^T_D[t]) = v(h^T_D)\]

\[(2)\quad s^p\text{ is a subgame-perfect equilibrium of } \mathcal{G}^\infty.\]

\[(3)\quad s_1^p \text{ and } s_2^p \text{ are computable.}\]

\[(4)\quad \text{For all } h \in \mathcal{H}^\infty, \text{ the continuation payoff satisfies:}\]

\[v(s^p; h) < \frac{1}{1 - \delta} v(h^T_D).\]

**Proof.** Assume that \( s \) is a Nash equilibrium satisfying the inequalities (a) and (b) for some \( \delta \in (0, 1) \). First, we show that there is a feasible and strictly individually rational payoff profile \( v^p \) of \( \mathcal{G}^\infty \) that is strictly Pareto dominated by \( v(s) \). Let \( v(s) = (1 - \delta)v(s) \) be the normalized payoff profile obtained by playing \( s \). Define \( m_1, m_2 \in A \) to be the minmax action profiles of the two players, and let \((x_1, y_1), (x_2, y_2)\) be their corresponding payoff profiles in \( G \):

\[x_1 = u_1(m_1) = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2),\]

\[x_2 = u_1(m_2),\]

\[y_1 = u_2(m_1),\]

\[y_2 = u_2(m_2) = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2).\]

Let \( \bar{v}^m \) be the minmax payoff profile in \( G \), that is:

\[\bar{v}^m = (x_1, y_2) = \left( \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2), \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2) \right).\]

Observe that \( \bar{v}^m \) might not be feasible.

Now, consider the 5 possible cases for the relationship between \((x_1, y_1)\) and \((x_2, y_2)\) as depicted in Figure 1:

1. \( x_1 = x_2 \) or \( y_1 = y_2 \). If \( x_1 = x_2 \), then \((x_1, y_2) = (x_2, y_2) = u(m_2)\) is a feasible payoff profile, and if \( y_1 = y_2 \), then \((x_1, y_2) = (x_1, y_1) = u(m_1)\) is a feasible payoff profile. Thus, in both cases \((x_1, y_2)\) is feasible, and by definition of \( x_1 \) and \( y_2 \), \((x_1, y_2)\) is the minmax payoff profile for both players. In this case, \( \bar{v}^m \) is trivially feasible. As \( v(s) \) is feasible by Theorem 2.19, then by the remarks after Definition 2.18, so is the normalized payoff profile \( \bar{v}^p = \frac{1}{2} v(s) + \frac{1}{2} \bar{v}^m \). By assumptions (a) and (b) we obtain \( v(s) = (1 - \delta)v(s) > \bar{v}^p \), and thus \( v(s) > \bar{v}^p \).

2. \( x_1 > x_2 \) and \( y_1 < y_2 \). Then, \( v(s) > (x_1, y_1) \) and \( v(s) > (x_2, y_2) \). There are coefficients \( \alpha_1, \alpha_2, \alpha_3 \in (0, 1) \) satisfying \( \alpha_1 + \alpha_2 + \alpha_3 = 1 \) and \( \alpha_1(x_1, y_1) + \alpha_2(x_2, y_2) + \alpha_3 v(s) = \bar{v}^m \), so \( \bar{v}^m \) is again feasible and by the same argument as in the previous case, there is a feasible and strictly individually rational payoff profile \( \bar{v}^p \).

3. \( x_1 < x_2 \) and \( y_1 > y_2 \). Write \( m_1 = (m_{1,1}, m_{1,2}) \) and \( m_2 = (m_{2,1}, m_{2,2}) \). By the definition of minmax, \( u(m_{2,1}, m_{1,2}) \leq (x_1, y_2) = \bar{v}^m \). Therefore there are coefficients \( \alpha_1, \alpha_2, \alpha_3 \in (0, 1) \) satisfying \( \alpha_1 + \alpha_2 + \alpha_3 = 1 \) and \( \alpha_1(x_1, y_1) + \alpha_2(x_2, y_2) + \alpha_3 u(m_{2,1}, m_{1,2}) = \bar{v}^m \), so \( \bar{v}^m \) is again feasible and by the same argument as before, there is a feasible and strictly individually rational payoff profile \( \bar{v}^p \).
Fig. 1. Possible relationships between minmax action profiles ($m_1$ and $m_2$). Horizontal (vertical) axis displays Player 1’s (Player 2’s) payoff, the highlighted area represents all individually rational payoff profiles, $s$ stands for the Nash equilibrium $s$. In the third case, $m_0$ is the action profile satisfying $u(m_0) \leq (x_1, y_2)$.

(4) $(x_1, y_1) < (x_2, y_2)$. Because $\tilde{v}(s) > \tilde{v}^m = (x_1, y_2)$, for any coefficients $\alpha_1, \alpha_2 \in (0, 1)$ satisfying $\alpha_1 + \alpha_2 = 1$ the payoff profile $\tilde{v}^p = \alpha_1 \tilde{v}(s) + \alpha_2 (x_1, y_1) < \tilde{v}(s)$ is feasible. Moreover, for large enough $\alpha_1$, $\tilde{v}^p > \tilde{v}^m$.

(5) $(x_1, y_1) > (x_2, y_2)$. Because $\tilde{v}(s) > \tilde{v}^m = (x_1, y_2)$, for any coefficients $\alpha_1, \alpha_2 \in (0, 1)$ satisfying $\alpha_1 + \alpha_2 = 1$ the payoff profile $\tilde{v}^p = \alpha_1 \tilde{v}(s) + \alpha_2 (x_2, y_2) < \tilde{v}(s)$ is feasible. Moreover, for large enough $\alpha_1$, $\tilde{v}^p > \tilde{v}^m$.

Thus, in all cases there is a feasible, normalized payoff profile $\tilde{v}^p$ satisfying $\tilde{v}^m < \tilde{v}^p < \tilde{v}(s)$. 

Now, define the normalized payoff profile
\[ \bar{\vartheta}^D = \frac{1}{2} \vartheta^p + \frac{1}{2} \vartheta(s) \]
and note that \( \bar{\vartheta}^D \) is feasible as it is a linear combination with non-negative coefficients of two feasible payoff profiles. Note that \( \vartheta^m < \vartheta^p < \bar{\vartheta}^D < \vartheta(s) \), and as all the normalized payoffs are normalized by \( 1 - \delta \), we have
\[ \vartheta^m < \vartheta^p < \bar{\vartheta}^D < \vartheta(s). \]
Define
\[ e^* = \frac{1}{3} \min_{i \in \{1,2\}} \min \left( \bar{\vartheta}^p_i - \vartheta^m_i, \bar{\vartheta}^D_i - \vartheta^p_i, \bar{\vartheta}^D_i - \vartheta(s)_i \right) \]
and observe that \( e^* > 0 \).

In the next step, we approximate the payoffs \( \bar{\vartheta}(s) \) and \( \bar{\vartheta}^D \) with finite histories \( h_C^T \) and \( h_D^T \) such that their length \( T \) is independent of the discount factor \( \delta \). As \( \bar{\vartheta}(s) \) and \( \bar{\vartheta}^D \) are both feasible, they can be written as linear combinations \( \bar{\vartheta}(s) = \sum_{a \in A} \alpha_a u(a) \) and \( \bar{\vartheta}^D = \sum_{a \in A} \beta_a u(a) \). For simplicity, let us assume that all coefficients \( \alpha_a, \beta_a \) are rational—if they are not, we can choose another feasible \( \bar{\vartheta}^C, \bar{\vartheta}^D \) from their neighbourhoods, still satisfying \( \bar{\vartheta}^m < \bar{\vartheta}^p < \bar{\vartheta}^D < \bar{\vartheta}(s) \). Then we can write
\[ \bar{\vartheta}(s) = \frac{1}{T} \sum_{i=1}^{T} u(h_C^T[i]) \]
and
\[ \bar{\vartheta}^D = \frac{1}{T} \sum_{i=1}^{T} u(h_D^T[i]). \]

We prove that for sufficiently large \( \delta \), the payoffs of \( h_C^T \) and \( h_D^T \) are both approximated by their normalized payoffs. The proof given below is for \( h_C^T \). The proof for \( h_D^T \) is identical, \textit{mutatis mutandis}.

Define
\[ \text{diff}(C) = \left| \frac{1}{T} \sum_{i=1}^{T} u(h_C^T[i]) - \frac{1 - \delta}{1 - \delta^T} \sum_{i=1}^{T} \delta^{i-1} u(h_C^T[i]) \right| \]
and
\[ \text{diff}(C) \leq \sum_{i=1}^{T} \left| \frac{1}{T} u(h_C^T[i]) - \frac{1 - \delta}{1 - \delta^T} \delta^{i-1} u(h_C^T[i]) \right| \leq \sum_{i=1}^{T} \left| \frac{1}{T} - \frac{1 - \delta}{1 - \delta^T} \delta^{i-1} \right| u(h_C^T[i]) \]
\[ \leq \left| \frac{1}{T} - \frac{1 - \delta}{1 - \delta^T} \delta^{l_{\text{max}}} \right| |u(h_C^T[i])| T \leq \left| \frac{1}{T} - \frac{1 - \delta}{1 - \delta^T} \delta^{l_{\text{max}}} \right| \bar{M}(G) T, \]
where
\[ \bar{M}(G) = \max_{a \in A, i \in \{1,2\}} |u_i(a)| \]
and
\[ i_{\text{max}} = \arg \max_{i \in \{1, \ldots, T\}} \left| \frac{1}{T} - \frac{1 - \delta}{1 - \delta_T} \delta^{i-1} - 1 \right|. \]

Now, define the constant
\[ \epsilon^* = \frac{\epsilon^*}{M(G)T}. \]

Observe that, for any \( i \in \{1, \ldots, T\} \) we have:
\[ \lim_{\delta \to 1} \frac{1 - \delta}{1 - \delta_T} \delta^{i-1} = \left( \lim_{\delta \to 1} \frac{1 - \delta}{1 - \delta_T} \right) \left( \lim_{\delta \to 1} \delta^{i-1} \right) = \left( \lim_{\delta \to 1} \frac{1}{\sum_{i=0}^{T-1} \delta^i} \right) \left( \lim_{\delta \to 1} \delta^{i-1} \right) = \frac{1}{T}. \]

Hence, there is a \( \delta^*_C < 1 \) such that for any \( \delta \) with \( \delta^*_C < \delta < 1 \) we have:
\[ \left| \frac{1}{T} - \frac{1 - \delta}{1 - \delta_T} \delta_{\text{max}} \right| < \epsilon^*, \]
and thus:
\[ \text{diff}(C) < \epsilon^*. \]

Similarly, for some \( \delta^*_D < 1 \) we have, for all \( \delta > \delta^*_D \) that:
\[ \text{diff}(D) = \left| \frac{1}{T} \sum_{i=1}^{T} u(h_D^T[i]) - \frac{1 - \delta}{1 - \delta_T} \sum_{i=1}^{T} \delta^{i-1} u(h_D^T[i]) \right| < \epsilon^*. \]

And thus, by definition of \( \epsilon^* \), we obtain \( v(h_D^T) > v(h_D^T) \) for all \( \delta > \max\{\delta^*_C, \delta^*_D\} \).

As \( \vartheta^p \) is feasible and strictly individually rational, by Theorem 2.19 with \( \epsilon = \epsilon^* \), there is a \( \delta^p < 1 \) such that for every \( \delta \) with \( \delta^p < \delta < 1 \) there is a subgame-perfect equilibrium \( s^p = (s_1^p, s_2^p) \) of \( G_\delta^{\infty} \), with \( s_1^p \) and \( s_2^p \) computable, leading to the normalized payoff profile \( \vartheta^p \). Moreover, by Theorem 2.19, the normalized continuation payoffs of \( s^p \) are within \( \epsilon^* \) of \( \vartheta^p \), hence we obtain \( v(s^p; h) < \frac{1}{1 - \delta_T} v(h_D^T) \).

Finally, setting \( \underline{\delta} = \max\{\delta^*_C, \delta^*_D, \delta^p\} \) concludes the proof.

**Theorem 3.4.** Let \( G \) be a 2-player normal-form game. If there exists a Nash equilibrium \( s' \) of \( G_\delta^{\infty} \) satisfying both
\[ v_1(s') > \frac{1}{1 - \delta} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) \]
and
\[ v_2(s') > \frac{1}{1 - \delta} \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2) \]
then there is a \( \delta \) with \( 0 < \delta < 1 \) such that for any \( \delta' \) with \( \delta < \delta' < 1 \), there is a strategy profile \( s = (s_1, s_2) \) of \( G_\delta^{\infty} \) satisfying:

1. \( s_2 \) is computable,
2. \( s_2 \) does not have a computable best response.

We split the proof of Theorem 3.4 into three separate parts—Lemma 3.14, Lemma 3.15, and Lemma 3.16, all below. Before we proceed to prove the lemmas let us define the essential structures needed to construct a subgame-perfect equilibrium with the desired properties.

\[ \text{Observe that if the function } f(x) = \frac{1}{x} - \frac{1 - \delta}{1 - \delta_T} \delta^x \text{ has the minimum in } x_{\text{min}}, f(x) \text{ is decreasing on } (-\infty, x_{\text{min}}) \text{ and increasing on } (x_{\text{min}}, \infty), \text{ whence } i_{\text{max}} \text{ is either 1 or } T, \text{ depending on the concrete values of } \delta \text{ and } T. \]
3.1 Definition of the strategy $\sigma_2$ and auxiliary notions

The high-level idea is to construct a prescribed path $\text{Path}_{\text{Pre}}$, and as long as $P1$ keeps playing the actions expected by the prescribed path, the strategy of $P2$ will not punish $P1$. There will also be two punishment paths, a mild punishment path $\text{Path}_{\text{MPun}}$ and a harsh punishment path $\text{Path}_{\text{HPun}}$. In $\text{Path}_{\text{MPun}}$, either the sequence $C_1$ and $C_2$ will be played repeatedly (roughly corresponding to cooperation in the Prisoner’s Dilemma), or the sequence $D_1$ and $D_2$ will be played (roughly corresponding to defection in the Prisoner’s Dilemma). In $\text{Path}_{\text{HPun}}$ we make use of a low-payoff equilibrium guaranteed to exist by the assumptions of Theorem 3.4.

The prescribed path of play is divided into finite segments of length $K = T+K_r+K_a$, consisting of a test period of length $T$, a reward period lasting $K_r = K_r'T$ stages, and an adjustment period lasting $K_a = K_a'T$ stages, where $K_r$ and $K_a$ are multiples of $T$. Thus, the strategy is parameterized in the length $T$ of the test period, the length $K_r$ of the reward period, and the factor $K_a$ that determines the length $K_a'T$ of the adjustment period. We shall later show that sagaciously choosing the triple $(T, K_r, K_a)$ together with a lower threshold $\delta$ for the discount factors suffice to prove our main result.

- Test periods are used to embed the recursively inseparable sets $A$ and $B$ problem into the strategies used: $P1$ is forced to play $C_1$ if the index $n$ of the current test period belongs to $A$ (or $D_1$ if $n \in B$), and $P2$ eventually checks if the action played is correct. Recursive enumerability of sets $A$ and $B$ ensures that the strategy of $P2$ is computable, while still allowing $P2$ to verify actions played by $P1$ in test periods. Recursive inseparability of sets $A$ and $B$ ensures that $P1$ has no computable best response. The payoffs from test periods are difficult to bound, and we make the test periods as short as possible so they have a little effect on the total payoff.
- Reward periods are the simplest: both players play high-payoff sequences $C_1$ and $C_2$ repeatedly, and the only purpose of reward periods is to ensure that the payoff on the prescribed path $\text{Path}_{\text{Pre}}$ is high enough.
- Adjustment periods serve to limit the difference between payoffs caused by deviations in test periods – necessary because there is no a priori upper bound on delay between a deviation and when it is detected – to ensure that the punishment is still effective. Both players play the high-payoff sequence $(C_1, C_2)$ if the payoff gained by $P1$ so far has been low (close to the minimum possible payoff when no player has deviated outside of test periods), and the low-payoff sequence $(D_1, D_2)$ if the payoff gained by $P2$ has been high. When adjustment periods are long enough, the payoff of $P1$ at the end of any adjustment period is independent of the sequence played in test periods, so from the payoffs’ point of view, all subgames starting at the first test stage are similar.

Lemma 3.3 provides a $T \in \mathbb{N}$, finite histories $h_C^T, h_D^T$ and a computable subgame-perfect equilibrium $s^p$. Thus, there are sequences of actions $C_1, C_2, D_1, D_2$ defined by:

$$C_1 = \left(h_C^T[i][1] \right)_{i=1}^T$$
$$C_2 = \left(h_C^T[i][2] \right)_{i=1}^T$$
$$D_1 = \left(h_D^T[i][1] \right)_{i=1}^T$$
$$D_2 = \left(h_D^T[i][2] \right)_{i=1}^T$$

That is, $C_1$ is the sequence of actions of $P1$ in $h_C^T$, $C_2$ is the sequence of actions of $P2$ in $h_C^T$, and so on.
Definition 3.5. Let \( G \) be a 2-player normal-form game. We define:

\[
\eta_G = \max_{a_1 \in A_1, a_2 \in A_2, i \in \{1, 2\}} u_i(a_1, a_2) - \min_{a_1 \in A_1, a_2 \in A_2, i \in \{1, 2\}} u_i(a_1, a_2).
\]

and, for a finite history \( h^T \in \mathcal{H}^T_{G^\infty} \), we define:

\[
\tilde{u}_i(h^T) = \frac{1}{\delta^T} \sum_{t=1}^{T} \delta^{T-t} u_i(h^T[t]).
\]

For each \( T \in \mathbb{N} \), we now define two subsets \( \mathcal{H}^T_{\text{test}}, \mathcal{H}^T_{\text{destest}} \subseteq \mathcal{H}^T_{G^\infty} \) of histories of length \( T \). The definitions use the sets \( \mathcal{A}_n \) and \( \mathcal{B}_n \) defined in Definition 2.24, and the notions of Test, Reward, and Adjustment stages (see Table 1).

The test set \( \mathcal{H}^T_{\text{test}} \) is the set of finite histories where the only allowed deviations from the prescribed path are at test periods not yet detected by P2. We define \( \mathcal{H}^T_{\text{test}} \), as well as two ancillary values \( u^T_{\text{test}} \) and \( \tilde{u}^T_{\text{test}} \), inductively on \( T \) below.

**Definition 3.6 (Test histories set).** \( \mathcal{H}^T_{\text{test}} \subseteq \mathcal{H}^T_{G^\infty} \) is defined inductively:

- \( \mathcal{H}^0_{\text{test}} = \mathcal{H}^0_{G^\infty} \) is the singleton set containing the empty history \( \lambda \) (of length 0) and \( u^0_{\text{test}} = \tilde{u}^0_{\text{test}} = 0 \).
- If \( h^t \in \mathcal{H}^T_{\text{test}} \), then \( T + 1 = (n - 1)(T + K_r + K_a) + t \) for some \( t, n \in \mathbb{N}, t \leq T \). Let \((a_1, a_2) = h^{T+1}[T + 1] \).
  - If \( T + 1 \) is a test stage. The history \( h^{T+1} \in \mathcal{H}^{T+1}_{\text{test}} \) if all of the following is satisfied:
    1. \( h^{T+1}_1[K(n - 1) + 1 : K(n - 1) + t] \leq C_1 \) or
    2. \( h^{T+1}_1[K(n - 1) + 1 : K(n - 1) + t] \leq D_1 \),
    3. for all \( m \in \mathbb{N}, m < n \) either:
      - \( h^{T+1}_1[K(m - 1) + 1 : K(m - 1) + T] = C_1 \) and \( m \in \mathcal{A}_n \), or
      - \( h^{T+1}_1[K(m - 1) + 1 : K(m - 1) + T] = D_1 \) and \( m \in \mathcal{B}_n \),
      - \( m \notin \mathcal{A}_n \cup \mathcal{B}_n \).
    Otherwise, \( h^{T+1} \notin \mathcal{H}^{T+1}_{\text{test}} \).
  - If \( T + 1 \) is a reward stage, then \( h^{T+1} \in \mathcal{H}^{T+1}_{\text{test}} \) if \((a_1, a_2) = (C_1[t], C_2[t]) \). Otherwise, \( h^{T+1} \notin \mathcal{H}^{T+1}_{\text{test}} \).

\[\begin{array}{|c|c|}
\hline
\text{Stage} & \text{Description} \\
\hline
nK + 1 & \text{Test period of length } T. \\
\vdots & \text{The sequence of actions played depends on the membership} \\
nK + T & \text{of the current stage in sets } \mathcal{A} \text{ and } \mathcal{B}. \\
\hline
nK + T + 1 & \text{Reward period of length } K_r. \\
\vdots & \text{Players play the high-payoff sequence of actions (C) to ensure} \\
nK + T + K_r & \text{that a unilateral deviation from the prescribed path is unprofitable.} \\
\hline
nK + T + K_r + 1 & \text{Adjustment period of length } K_a. \\
\vdots & \text{High-payoff sequence (C) if the discounted payoff of P1} \\
nK + T + K_r + K_a & \text{so far has been low, low-payoff sequence (D) otherwise.} \\
\hline
\end{array}\]

Table 1. Every \( K \) stages of the prescribed path are split into test, reward and adjustment period.
If $T + 1$ is an adjustment stage, then $h^{T+1} \in \mathcal{H}^{T+1}_{test}$ if either:

* $(a_1, a_2) = (C_1[t], C_2[t])$ and $\bar{u}_t(h^T[1 : T - t]) - u^{T-t}_{test} \leq \eta_G$, or
* $(a_1, a_2) = (D_1[t], D_2[t])$ and $\bar{u}_t(h^T[1 : T - t]) - u^{T-t}_{test} > \eta_G$.

Otherwise, $h^{T+1} \not\in \mathcal{H}^{T+1}_{test}$.

Furthermore, we define:

\[
\bar{u}^T_{test} = \min_{h^T \in \mathcal{H}^T_{test}} \bar{u}_1(h^T),
\]
\[
\bar{u}^T_{test} = \max_{h^T \in \mathcal{H}^T_{test}} \bar{u}_1(h^T).
\]

Observe that both $\eta_G \geq 0$ and $\bar{u}^T_{test} - u^T_{test} \geq 0$ (even if some payoffs are negative).

**Proposition 3.7.** Let $T, m \geq 1$ be integers. Then:

\[
\delta^m(\bar{u}^{T+m}_{test} - u^{T+m}_{test}) - (\bar{u}^T_{test} - u^T_{test}) \leq \delta^m \left( \min_{a_1, a_2} \max_{a_1, a_2} u_1(a_1, a_2) - \min_{a_1, a_2} u_1(a_1, a_2) \right)
\]

\[
\leq \frac{1}{\delta^T} \sum_{t = T+1}^{T+m} (\delta^{t-1} \eta_G) = \sum_{t = T+1}^{T+m} \delta^{t-1} \eta_G \leq m \eta_G
\]

**Proof.** Observe that:

\[
\delta^m(\bar{u}^{T+m}_{test} - u^{T+m}_{test}) - (\bar{u}^T_{test} - u^T_{test}) \leq \delta^m \left( \min_{a_1, a_2} \max_{a_1, a_2} u_1(a_1, a_2) - \min_{a_1, a_2} u_1(a_1, a_2) \right)
\]

\[
\leq \frac{1}{\delta^T} \sum_{t = T+1}^{T+m} (\delta^{t-1} \eta_G) = \sum_{t = T+1}^{T+m} \delta^{t-1} \eta_G \leq m \eta_G
\]

Proposition 3.7 does not use the fact that test histories have strict requirements – in most stages, there is only one allowed action in each stage, independent of the previously played actions. With this stronger assumption, we get the following result.

**Proposition 3.8.** Let $T, m \geq 1$ be integers and assume that all test histories $h \in \mathcal{H}^{T+m}_{test}$ share the subsequence $h[T+1 : T+m]$. Then:

\[
\delta^m(\bar{u}^{T+m}_{test} - u^{T+m}_{test}) - (\bar{u}^T_{test} - u^T_{test}) \leq \frac{1}{\delta^T} \sum_{t = T+1}^{T+m} (\delta^{t-1} \eta_G) = \sum_{t = T+1}^{T+m} \delta^{t-1} \eta_G \leq m \eta_G
\]

**Proof.** Let $h \in \mathcal{H}^{T+m}_{test}$ be any test history. Observe that:

\[
\delta^m(\bar{u}^{T+m}_{test} - u^{T+m}_{test}) - (\bar{u}^T_{test} - u^T_{test}) \leq \delta^m \left( \frac{1}{\delta^{T+m}} \sum_{t = T+1}^{T+m} (\delta^{t-1} \eta_G) \right) = 0
\]

The set of finite histories $\mathcal{H}_{devtest}$ is a subset of $\mathcal{H}^{\infty}_{G}$ where the only allowed deviations are deviations by P1 at test periods, and if such a deviation is detected, play will switch to the punishment path $\text{Path}_{\text{Pun}}$ repeating the sequence $(h^T_D)^3 h^T_C$ (i.e., 3 copies of $h^T_D$, followed by a single copy of $h^T_C$), of length $4T$ forever.

**Definition 3.9 (Deviation test set).** The $\mathcal{H}_{devtest}$ is defined by induction as follows:

- $\mathcal{H}_{devtest}^0 = \mathcal{H}^{\infty}_{G}$ consists of the only (empty) history of length 0.
- If $h^{T+1} \in \mathcal{H}_{test}^{T}$, then $h^{T+1} \in \mathcal{H}_{devtest}^{T+1}$.
- If $h^{T+1} \not\in \mathcal{H}_{test}^{T+1}$, then $T + 1 = (n - 1)(T + K_r + K_a) + t$ for some $t, n \in \mathbb{N}, t \leq T$. Let $(a_1, a_2) = h^{T+1}[T + 1]$. If $h^{T+1-t} \in \mathcal{H}_{test}^{T+1-t}$, we consider two cases to determine whether $h^{T+1} \in \mathcal{H}_{devtest}^{T+1}$.
* If $T + 1$ is a test stage, i.e. the stage in the test period when P2 detects a deviation by P1 in a test period before $T + 1 - t = (n - 1)K$. Then the history $h_{T+1}^T \in H_{test}^{T+1}$ if and only if $h_{1}^{T+1}[(n-1)K+1:T+1] \leq D_1$ and $h_{2}^{T+1}[(n-1)K+1:T+1] \leq D_2$.

* If $T + 1$ is a reward or adjustment stage, then $h_{T+1}^T \notin H_{de/test}^{T+1}$.

- Else, if $h^T \in H_{de/test}^{T+1}$, let $T'$ be the greatest integer such that $T' < T$ and $h^T[1:T] \in H_{test}^{T'}$, but $h^T[1:T'+1] \notin H_{test}^{T+1}$. Because such test deviations are detected in the beginning of test periods, $T|T'$. Let $\Delta = (T + 1 - t - T')/T$. Divisibility of $\Delta$ by 4 determines whether the next move is $h_{1}^T[t]$ or $h_{2}^T[t]$:

  * If $4|\Delta$ and $(a_1, a_2) = (C_1[t], C_2[t])$, then $h_{T+1}^T \in H_{de/test}^{T+1}$.
  * If $4 \nmid \Delta$ and $(a_1, a_2) = (D_1[t], D_2[t])$, then $h_{T+1}^T \in H_{de/test}^{T+1}$.
  - Otherwise, $h_{T+1}^T \notin H_{de/test}^{T+1}$.

We proceed by defining the strategy profile in $G^\infty$ and later we will prove that it satisfies the conditions of Theorem 3.4.

As with the test sets, both strategies in the profile are parameterized over $T$, $K_r$, and $K_a$.

From Lemma 3.3, we can assume that there is a subgame-perfect equilibrium $s^p$ satisfying for any finite history $h^T$:

$$v_i(s^p; h^T) < \frac{1}{1 - \delta^T} v_i(D_1, D_2) < \frac{1}{1 - \delta^T} v_i(C_1, C_2) \forall i \in \{1, 2\}.$$

**Definition 3.10 (The strategy $s_2$).** Let $s_2$ be the strategy of P2 that plays the following, given a finite history $h^T \in H_{G}^\infty$ with $T + 1 = (n - 1)K + t$ and $1 \leq t \leq K$:

1. If $h^T \in H_{test}^{T+1}$, consider two cases:
   - If $T + 1$ is a test or reward stage, play $C_2[t]$.
   - If $T + 1$ is an adjustment stage, play $C_2[t]$ if $\bar{u}_i(h^T) - u_{test}^i \leq \eta G$, play $D_2[t]$ otherwise.

2. Else, if $h^T \in H_{de/test}^{T+1}$, let $T'$ be the greatest integer such that $T' < T$ and $h^T[1:T'] \in H_{test}^{T'}$, but $h^T[1:T'+1] \notin H_{test}^{T+1}$. Let $\Delta = (T + 1 - t - T')/T$. Play $C_2[t]$ if $4|\Delta$, otherwise play $D_2[t]$.

3. Otherwise play $s_2^p(h^T)$.

As the punishment path Path$_{MPun}$ consists of repeating the sequence $(h_D^T)^3h_C^T$ of length $4T$, the condition $4|\Delta$ is used to determine whether the next action played is $D_2[t]$ or $C_2[t]$ (analogously, $D_1[t]$ or $C_1[t]$ for Player 1). To ensure the subgame-perfect equilibrium, P1 plays the optimal action in any test stage following a history that contains a mild deviation by P1 in a test stage. Observe that the optimal action in this case is still either $C_1[t]$ or $D_1[t]$.

**Definition 3.11 (The strategy $s_1$).** Let $s_1$ be the strategy of P1 that plays the following, given a finite history $h^T \in H_{G}^\infty$ with $T + 1 = (n - 1)K + t$ and $1 \leq t \leq K$:

1. If $h^T \in H_{test}^{T+1}$, consider three cases:
   - If $T + 1$ is a test stage. If P1 has not deviated in a test stage before, play $C_1[t]$ if $n \in A$, play $D_1[t]$ if $n \in B$. Otherwise play the optimal action.
   - If $T + 1$ is a reward stage, play $C_1[t]$.
   - If $T + 1$ is an adjustment stage, play $C_1[t]$ if $\bar{u}_i(h^T) - u_{test}^i \leq \eta G$, play $D_1[t]$ otherwise.

2. Else, if $h^T \in H_{de/test}^{T+1}$, let $T'$ be the greatest integer such that $T' < T$ and $h^T[1:T'] \in H_{test}^{T'}$, but $h^T[1:T'+1] \notin H_{test}^{T+1}$. Let $\Delta = (T + 1 - t - T')/T$. Play $C_1[t]$ if $4|\Delta$, otherwise play $D_1[t]$.

3. Otherwise play $s_1^p(h^T)$.
3.2 Some auxiliary lemmas

We now proceed to prove a sequence of lemmas necessary to prove Theorem 3.4.

**Lemma 3.12.** For any \( \delta \), there is a constant \( u_\delta > 0 \) such that the difference

\[
\hat{u}_{\text{test}}^{T+i} - u_{\text{test}}^{T+i}
\]

decreases at least by \( u_\delta \) every \( T \) adjustment stages, provided that \( \eta_G < \hat{u}_{\text{test}}^{T+i} - u_{\text{test}}^{T+i} < 3T \eta_G \) and that \( T+1 \) is the first stage of an adjustment period.

**Proof.** To prove this, let \( i \in \{0, T, 2T, \ldots, (K'_a - 1)T\} \) be given and consider \( j = T + i \). Note that the prescribed play in adjustment stages is either the sequence \((D_1, D_2)\) or \((C_1, C_2)\). It follows that \( \hat{u}_{\text{test}}^{j+i} \leq \frac{1}{\delta^T}(\hat{u}_{\text{test}}^j + v_1(C_1, C_2)) \) and \( u_{\text{test}}^{j+i} \geq \frac{1}{\delta^T}(u_{\text{test}}^j + v_1(D_1, D_2)) \).

For a given \( \delta \) define

\[
u_\delta = \frac{1}{\delta^T} \left( v_1(C_1, C_2) - v_1(D_1, D_2) - 3T \eta_G (1 - \delta^T) \right)\]

Then we have:

\[
(u_{\text{test}}^{j+i} - u_{\text{test}}^j) - (u_{\text{test}}^{j+i} - u_{\text{test}}^j) \geq (u_{\text{test}}^{j+i} - u_{\text{test}}^j) - \frac{1}{\delta^T} (u_{\text{test}}^{j+i} - u_{\text{test}}^j) - \frac{1}{\delta^T} (v_1(C_1, C_2) - v_1(D_1, D_2))
\]

\[
= \frac{1}{\delta^T} \left( v_1(C_1, C_2) - v_1(D_1, D_2) - (u_{\text{test}}^j - u_{\text{test}}^j)(1 - \delta^T) \right)
\]

\[
> \frac{1}{\delta^T} \left( v_1(C_1, C_2) - v_1(D_1, D_2) - 3T \eta_G (1 - \delta^T) \right)
\]

\[
= u_\delta.
\]

Observe that \( u_\delta > 0 \) if

\[
\delta^T > 1 - \frac{v_1(C_1, C_2) - v_1(D_1, D_2)}{3T \eta_G}
\]

and note that

\[
1 - \frac{v_1(C_1, C_2) - v_1(D_1, D_2)}{3T \eta_G} \geq 1 - \frac{1}{3} > 0
\]

\( \Box \)

The following lemma states that the difference between Player 1’s payoffs in any two histories from \( H^T_{\text{test}} \) is bounded. Later, we will show that for large enough discount factor, this bound is smaller than the possible loss from deviating from the prescribed path.

**Lemma 3.13.** For any \( K_a \in \mathbb{N} \), there are (i) \( K_a \in \mathbb{N} \), and (ii) \( \delta \) with \( 0 < \delta < 1 \) such that for any discount factor \( \delta \) with \( \delta < \delta < 1 \), any \( T \in \mathbb{N} \), and any \( h_1^T, h_2^T \in H^T_{\text{test}} \):

\[
|\hat{u}_1(h_1^T) - \hat{u}_1(h_2^T)| < 3T \eta_G.
\]

**Proof.** First observe that \( |\hat{u}_1(h_1^T) - \hat{u}_1(h_2^T)| \leq \hat{u}_{\text{test}}^T - u_{\text{test}}^T \). We shall prove the following two inequalities by induction on \( T \):

**Weak inequality:** \( \hat{u}_{\text{test}}^T - u_{\text{test}}^T < 3T \eta_G \)

**Strong inequality:** If \( T = 1 \) or \( T + 1 \) is the first stage of a test period, \( \hat{u}_{\text{test}}^T - u_{\text{test}}^T < \frac{1}{\delta^T} \eta_G \)

Note that the lemma follows from the weak inequality; the strong inequality is needed in the induction to prove the weak inequality.
Base step: If $T = 1$, note first that $T$ is the first stage of a test period. We have:

\[
\tilde{u}_{T_{\text{test}}}^T - \tilde{u}_{T_{\text{test}}}^T = \max_{h' \in \mathcal{H}_{\text{test}}} \tilde{u}_1(h^1) - \min_{h' \in \mathcal{H}_{\text{test}}} \tilde{u}_1(h^1) \leq \frac{1}{\delta} \left( \max_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2) - \min_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2) \right)
\]

\[
\leq \frac{1}{\delta} \left( \max_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2) - \min_{a_1 \in A_1, a_2 \in A_2, i \in \{1,2\}} u_1(a_1, a_2) \right) = \frac{1}{\delta} \eta_G
\]

Thus, the strong inequality is satisfied, and choosing any $\delta \geq 1/3T$ suffices to satisfy the weak inequality.

Induction step: Assume that the induction hypothesis holds for $T$, where $T + 1$ is either a test, reward, or adjustment stage.

T+1 is a test stage. Then, there is a $T' < T + 1$ such that the strong inequality holds for $T'$ (either $T' + 1$ is the first stage of a test period, or $T' = 1$), and such that stages $T' + 1, \ldots, T' + T$ are test stages, and there is some $i \in \{1, \ldots, T\}$ with $T' + i = T + 1$. By Proposition 3.7 and the induction hypothesis for $T'$ we obtain:

\[
\tilde{u}_{T_{\text{test}}}^{T' + i} - \tilde{u}_{T_{\text{test}}}^{T' + i} \leq \frac{1}{\delta T'} \left(i \eta_G + \frac{\eta_G}{\delta T'} \right) = \frac{\eta_G}{\delta T'} \left(i + \frac{1}{\delta T'} \right) \leq \frac{\eta_G}{\delta T'} \left(T' + \frac{1}{\delta T'} \right)
\]

Now, choosing $\delta$ large enough that any $\delta > \delta$ satisfies $(T' + 1/\delta T')/\delta T' \leq 3T'$ satisfies the weak inequality. As $T + 1$ is a test stage, $T + 2$ is not the first stage of a test period, and there is nothing to prove for the strong inequality.

T+1 is a reward stage. Then there is $T' < T + 1$ such that (i) either $T' = 1$ or $T' + 1$ is the first stage of a test period, and (ii) stages $T' + T + 1, \ldots, T' + T + K_r$ are reward stages, and (iii) there is some $i \in \{1, \ldots, K_r\}$ such that $T + 1 = T' + T + i$.

By Proposition 3.8 and the inequality (5) applied to the test stage $T' + T$, we have:

\[
\tilde{u}_{T_{\text{test}}}^{T' + T + i} - \tilde{u}_{T_{\text{test}}}^{T' + T + i} \leq \frac{1}{\delta (T + 1)} \left(\tilde{u}_{T_{\text{test}}}^{T' + T} - \tilde{u}_{T_{\text{test}}}^{T' + T} \right) < \frac{\eta_G}{\delta T} \left(T' + 1/\delta T \right) \leq \frac{\eta_G}{\delta T + K_r} \left(T' + 1/\delta T \right)
\]

Clearly, for all large enough $\delta$:

\[
\frac{\eta_G}{\delta T + K_r} \left(T' + 1/\delta T \right) < 2T \eta_G
\]

and there is thus some $\overline{\delta}$ such that for all $\delta < \overline{\delta}$:

\[
\tilde{u}_{T_{\text{test}}}^{T+1} - \tilde{u}_{T_{\text{test}}}^{T+1} < 2T \eta_G < 3T \eta_G.
\]

and hence the weak inequality holds. As $T + 2$ is not a test stage (because a reward stage is followed either by another reward stage or an adjustment stage), there is nothing to prove for the strong inequality.

T+1 is an adjustment stage. Then there is $T' < T$ such that either $T' = 1$, or $T' + 1$ is a test stage, stages $T' + T + K_r + 1, \ldots, T' + K$ are adjustment stages, and there is some $i \in \{1, \ldots, K_a\}$ such that $T + 1 = T' + T + K_r + i$.

By the inequality (6) applied to the reward stage $T' + T + K_r$:

\[
\tilde{u}_{T_{\text{test}}}^{T' + T + K_r} - \tilde{u}_{T_{\text{test}}}^{T' + T + K_r} < 2T \eta_G
\]

and by Lemma 3.12, this difference decreases by $\tilde{u}^3 > 0$ every $T$ adjustment stages. Hence, for all $i \in \{T, 2T, \ldots, K_a\}$ we have:

\[
\tilde{u}_{T_{\text{test}}}^{T' + T + K_r + i} - \tilde{u}_{T_{\text{test}}}^{T' + T + K_r + i} < 2T \eta_G
\]
For increasing \( i \notin \{ T, 2T, \ldots, K_i T \} \), the difference may increase, but because there is only one allowed action in any stage, by Proposition 3.8 we have \( (\bar{u}_{test}^{T+T+K_i T+i} - u_{test}^{T+T+K_i T+i}) \leq \frac{1}{\delta} (\bar{u}_{test}^{T+T+K_i T+i-1} - u_{test}^{T+T+K_i T+i-1}) \). Hence, for all \( i \in \{1, \ldots, K_a\} \):

\[
\bar{u}_{test}^{T+T+K_i T+i} - u_{test}^{T+T+K_i T+i} < \frac{1}{\delta^T} 2^T \eta_G
\]

and for all sufficiently large \( \delta \):

\[
\bar{u}_{test}^{T+T+K_i T+i} - u_{test}^{T+T+K_i T+i} < 3^T \eta_G
\]

thus proving the weak inequality. For all stages \( T \in \{ T' + T + K_r + 1, \ldots, T' + T + K_r + K - 1 \} \), \( T + 1 \) is not a test stage, whence there is nothing to prove for the strong inequality. However, for stage \( T' + K = T' + T + K_r + K_a \), stage \( T + K + 1 \) is a test stage, and the strong inequality must be proved. Thus, consider stage \( T + K \), and observe that if \( K_a \) satisfies

\[
K_a \geq \frac{3^T \eta_G}{u_{\delta}}
\]

then as each multiple of \( T' \) stages in the adjustment period decreases the difference \( \bar{u}_{test}^{T+T+K_i T+i} - u_{test}^{T+T+K_i T+i} \) by \( u_{\delta}^T \) (while \( \eta_G < u_{test}^{T+K_r T+i} - u_{test}^{T+K_r T+i+1} \)), there is \( l \in \{0, T, 2T, \ldots, K_a = K_a' T \} \) such that \( u_{test}^{T+T+K_r T+i} - u_{test}^{T+T+K_r T+i+1} < \eta_G \). Each adjustment stage has a single allowed action, hence for all integers \( j \) with \( l \leq j \leq K_a \):

\[
\bar{u}_{test}^{T+K_r T+i+j} - u_{test}^{T+K_r T+i+j} < \frac{\eta_G}{\delta^T}
\]

This finishes the proof by induction, and it thus remains to set \( \delta \) and \( K_a \). Define \( \delta \) to be the least real number that (i) satisfies the values for \( \delta \) in the base step and the three cases of the induction step, including inequality (4), and (ii) is yielded by Lemma 3.3. Moreover, for this \( \delta \), we set

\[
u_\delta = \frac{1}{\delta^T} \left( v_1(C_1, C_2) - v_1(D_1, D_2) - 3^T \eta_G (1 - \delta^T) \right)
\]

and

\[
K_a = T \left[ \frac{3 \eta_G}{u_\delta} \right]
\]

so that \( K_a \) is independent of the discount factor \( \delta \).

We next prove that \((\sigma_1, \sigma_2)\) is a subgame-perfect equilibrium.

**Lemma 3.14.** There are \( K_a, K_r \in \mathbb{N} \) and \( \delta \) with \( 0 < \delta < 1 \) such that for any discount factor \( \delta \) satisfying \( \bar{\delta} \leq \delta < 1 \), the strategy profile \( s = (\sigma_1, \sigma_2) \) is a subgame-perfect equilibrium.

**Proof.** Let \( K_a \) be the value provided by Lemma 3.13. We start by computing the payoff on the prescribed path \( \text{Path}_{pre} \). Let \( v \) be the least possible payoff of any player at the test period, that is:

\[
v = \min \{ v_1(C_1, C_2), v_1(D_1, C_2), v_2(C_1, C_2), v_2(D_1, C_2) \}
\]

The discounted payoff of \( Pi \) in the first \( K \) stages is at least

\[
\bar{u}_{i}^{K} \text{ (Path}_{pre} \text{)} \geq v + \delta^T \frac{1 - \delta^K_r}{1 - \delta^T} v_1(C_1, C_2) + \delta^{K_r + T} \frac{1 - \delta^K_a}{1 - \delta^T} v_1(D_1, D_2)
\]

Set \( y_i = v (1 - \delta^T)/v_1(D_1, D_2) \), \( \alpha_i^C = \delta^T (1 - \delta^K_r) \), and \( \alpha_i^D = y_i + \delta^{K_r + T} (1 - \delta^K_a) \). Then:

\[
\bar{u}_{i}^{K} \text{ (Path}_{pre} \text{)} (1 - \delta^T) \geq \alpha_i^C v_1(C_1, C_2) + \alpha_i^D v_1(D_1, D_2)
\]
Observe that $\alpha^C$ increases with increasing $\delta$ and $K_r$, and that $\alpha^D$ decreases with increasing $K_r$. Because $K_a$ is independent of $\delta$, for all large enough $\delta$ and $K_r$, we have:

$$\tilde{u}_i^K(\text{Path}_{\text{Pre}}) \geq \frac{1}{2(1 - \delta^T)}(v_i(C_1, C_2) + v_i(D_1, D_2)),$$

and thus, the present value of the payoff of $P_i$ at the beginning of any test period satisfies:

$$\tilde{u}_i(\text{Path}_{\text{Pre}}) = \frac{\tilde{u}_i^K(\text{Path}_{\text{Pre}})}{1 - \delta^K} > \frac{1}{2(1 - \delta^T)}(v_i(C_1, C_2) + v_i(D_1, D_2))$$

The mild punishment path $\text{Path}_{\text{MPun}}$ starts at stage $T$ when $P_2$ first detects a deviation by $P_1$ at an earlier test stage, and consists of repeated play of the sequence $(D_1, D_2)^3(C_1, C_2)$ of length $4T$. As $P_2$ detects such a deviation at the first stage of a test period, $T - 1$ is an adjustment stage. The discounted payoff of $P_i$ after the first $4T$ stages of $\text{Path}_{\text{MPun}}$ is:

$$\frac{1 - \delta^{3T}}{1 - \delta^T}v_i(D_1, D_2) + \delta^{3T}v_i(C_1, C_2),$$

so the discounted value of the payoff of $P_i$ at stage $T$ is:

$$\tilde{u}_i(\text{Path}_{\text{MPun}}) = \frac{\delta^{3T}}{1 - \delta^T}v_i(C_1, C_2) + \frac{1 - \delta^{3T}}{(1 - \delta^T)(1 - \delta^T)}v_i(D_1, D_2)$$

The harsh punishment path $\text{Path}_{\text{HPun}}$ begins at stage $T$ when any player deviates from a history $h^{T-1} \in H_{\text{test}} \cup H_{\text{deatest}}$, except for a mild deviation by $P_1$ at a test stage (which leads to $\text{Path}_{\text{MPun}}$). $\text{Path}_{\text{HPun}}$ is the equilibrium path induced by $s^p$, so by Lemma 3.3 the present value of the payoff of $P_1$ at stage $T$ is:

$$\tilde{u}_i(\text{Path}_{\text{HPun}}) = v_i(s^p; h^{T-1}) < \frac{1}{1 - \delta^T}v_i(D_1, D_2)$$

To prove that $s$ is a subgame-perfect equilibrium, we apply the One-Shot Deviation Principle (Theorem 2.17) and argue that any one-shot deviation from $s$, following any finite history, would result in a lower payoff for both players. There are four cases of deviation:

- **P1 deviates from $\text{Path}_{\text{Pre}}$ at a test stage $T = nK + t_0$ with $n \geq 0$ and $1 \leq t_0 \leq T$, and $P_2$ first detects this kind of deviation at a test stage $n'K + 1$ for some $n' > n$, the game then switches to $\text{Path}_{\text{MPun}}$. Note that P1 can now play optimally at test stages, as long as the play leads to a history $h^{n'K} \in H_{\text{test}}$. By Lemma 3.13, for any other history $h' \in H_{\text{test}}^{n'K}$ following the prescribed path, we have $\tilde{u}_1(h^{n'K}) - \tilde{u}_1(h') < 3T \eta G$. Now observe that (8) implies that:

$$\lim_{\delta \to 1} \tilde{u}_1(\text{Path}_{\text{MPun}}) = \frac{1}{4(1 - \delta^T)}v_1(C_1, C_2) + \frac{3}{4(1 - \delta^T)}v_1(D_1, D_2)$$

Hence, for $\delta$ large enough, the difference of the discounted payoff for $P_1$ at the beginning of stage $n'K + 1$ is $\tilde{u}_1(\text{Path}_{\text{Pre}}) - \tilde{u}_1(\text{Path}_{\text{MPun}}) > 3T \eta G$, so the Player 1’s payoff decreases by deviating.

- **Any player $P_i$ deviates from $\text{Path}_{\text{Pre}}$ at a stage $T$ and the game switches to $\text{Path}_{\text{HPun}}$ at stage $T + 1$. If $P_i$ deviates, the maximum possible discounted payoff is bounded above by $\tilde{w}_i = \bar{u} + \tilde{u}_i(\text{Path}_{\text{HPun}})$, where $\bar{u} = \max_{a_1 \in A_1, a_2 \in A_2, i \in \{1, 2\}} u_i(a_1, a_2)$. We now compute a lower bound $w_i$ on the payoff for $P_i$ if this deviation does not happen (even if $P_2$ later detects an earlier deviation by $P_1$ at a test stage and the game switches to $\text{Path}_{\text{MPun}}$). In the worst case – when no sequence $(C_1, C_2)$ is played before switching to the mild punishment path – players finish the current block of length $T$ (with no guarantee on the payoff for the first $T$ stages) and...
subsequently both play the low-payoff sequence \((D_1, D_2)\) \(K_a\) times, after which, in the worst case, the game switches to Path_{MPun}. Thus:

\[
\bar{w}_i = \frac{1 - \delta^T}{1 - \delta} u + \delta^T \frac{1 - \delta} {1 - \delta^T} v_i(D_1, D_2) + \delta^K \hat{u}_i(\text{Path}_{MPun}),
\]

where \(u = \min_{a_i \in A, a_j \in A, i \in \{1,2\}} u_i(a_1, a_2).\) For large enough \(\delta, \bar{w}_i < \bar{w}_j,\) so the maximum possible gain of \(P_i\) from deviating is again negative.

- Any player deviates from Path_{MPun} at a stage \(T\) and the game switches to Path_{HPun} at stage \(T + 1\). The maximum possible gain from deviating in the first \(T\) stages is bounded above by \(T \eta_G,\) followed by payoff \(\delta^T \hat{u}_i(\text{Path}_{HPun})\) in case of a deviation and followed by payoff \(\delta^T \hat{u}_i(\text{Path}_{MPun})\) in case no deviation happened. But for all sufficiently large enough \(\delta,\)

\[
\delta^T \hat{u}_i(\text{Path}_{MPun}) - \delta^T \hat{u}_i(\text{Path}_{HPun}) > T \eta_G,\n\]

whence the deviating player’s payoff decreases.

- Any player deviates from Path_{HPun}. By definition of \(s^p,\) Path_{HPun} is a subgame-perfect equilibrium, and thus this deviation is unprofitable for any player.

Define \(\delta\) to be the least \(\delta\) yielded by Lemma 3.13 and satisfying inequality (7), \(\hat{u}_1(\text{Path}_{pre}) - \hat{u}_1(\text{Path}_{MPun}) > 3T \eta_G,\) \(\bar{w}_i < \bar{w}_j\) and \(\delta^T \hat{u}_i(\text{Path}_{MPun}) - \delta^T \hat{u}_i(\text{Path}_{HPun}) > T \eta_G.\) Then, \(\delta < 1,\) and \(s\) is a subgame-perfect equilibrium, as desired.

As expected, \(\sigma_2\) is computable:

**Lemma 3.15.** The strategy \(\sigma_2\) is computable.

**Proof.** The game has three possible states - Path_{pre}, Path_{MPun}, and Path_{HPun}. In the state Path_{MPun}, \(\sigma_2\) repeats a finite sequence of actions and checks whether \(P_1\) played the prescribed action in the previous stage—which is also easily computed from a finite sequence of actions and the current stage number. In the state Path_{HPun}, \(\sigma_2\) uses the computable strategy \(s_2^p\) to compute the next move. Thus, the only potentially problematic case is the prescribed path Path_{pre}. If the current stage \(T + 1 = (n - 1)(T + K_r + K_a) + t\) is a test stage, \(P_2\) uses a universal Turing machine to simulate one more step of each of computations of \(\phi_1(1), \ldots, \phi_n(n)\) (using encodings of Turing machines \(T_1, \ldots, T_n\)), and if any of the computations halt, \(P_2\) verifies whether \(P_2\) played the prescribed sequence of actions in the corresponding test period. In case of a deviation in a test period, the game switches to Path_{MPun}. The next action played is simply \(C_2[t]\) if the current stage is a test or reward stage. To determine the next action in an adjustment stage, \(P_2\) must decide whether \(\hat{u}_1(h^T) - u^T_{\text{test}} \leq \eta_G.\) Here, \(\eta_G\) is a constant, and the current stage, \(\hat{u}_1(h^T)\) is the payoff of \(P_1\) obtained so far. To compute \(u^T_{\text{test}},\) \(P_2\) only updates the actions played by \(P_1\) in the test stages of \(h^T\) where the corresponding Turing machines have not yet halted, and then computes the payoff for the updated history. Lastly, \(P_2\) verifies whether the game should switch to Path_{HPun}; note that this happens if \(P_1\) played an incorrect action in the previous stage (except for a mild deviation at a test stage). Again, this is straightforward if the previous stage was a test or a reward stage. If the previous stage was an adjustment stage, \(P_2\) determines the prescribed action in the previous stage again based on the inequality \(\hat{u}_1(h^{T-1}) - u^{T-1}_{\text{test}} \leq \eta_G.\)

And, finally, we have:

**Lemma 3.16.** No best response to \(\sigma_2\) is computable.

**Proof.** Let \(s_1\) be a best response to \(\sigma_2.\) \(P_1\) must play the prescribed action in every test stage, or the game will eventually switch to one of the punishment paths, resulting in strictly less payoff for \(P_1\) compared to staying on the prescribed path (see the proof of Lemma 3.14). For a test stage \(T = K(n - 1) + t, h_1^n[T] = C_1[t]\) if \(n \in \mathcal{A}\) and \(h_1^n[T] = D_1[t]\) if \(n \in \mathcal{B}.\) As \(s_1\) is computable, there
is a Turing machine $T_{s_1}$ computing $s_1$. However, the Turing machine that simulates $T_{s_1}$, but accepts whenever $C_1\{t\}$ is output and rejects whenever $D_1\{t\}$ is output, decides a language $C$ that separates $\mathcal{A}$ and $\mathcal{B}$, thus contradicting Proposition 2.23.

3.3 A special case: shallow games

The last class of games we have not considered to obtain a complete characterization are the games where every payoff of P2 under a Nash equilibrium of $G_\delta^\infty$ equals their minmax payoff in the stage game $G$. We call such games shallow games.

Definition 3.17. Let $G$ be a 2-player normal-form game. $G$ is said to be shallow if the following two conditions are both satisfied:

1. There is a $\delta$ with $0 < \delta < 1$ and a Nash equilibrium $s'$ of $G_\delta^\infty$ that satisfies:
   \[
   v_1(s') > \frac{1}{1 - \delta} \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2).
   \]

2. For every $\delta$ with $0 < \delta < 1$, every Nash equilibrium $s''$ of $G_\delta^\infty$ satisfies:
   \[
   v_2(s'') = \frac{1}{1 - \delta} \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2).
   \]

Proposition 3.18. Let $G$ be shallow. Every action profile $a \in A$ satisfies
\[
   u_2(a) \leq \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2).
\]

Proof. Assume, for contradiction, that there is an action profile $a \in A$ with
\[
   u_2(a) > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2).
\]

For any $\epsilon$ with $0 < \epsilon < 1$ the normalized payoff profile $\tilde{\nu}^* = (1 - \epsilon)\tilde{\nu}(s') + \epsilon u(a)$ is feasible and satisfies $\tilde{\nu}_2^* > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)$. Moreover, for small enough $\epsilon$, the condition $\tilde{\nu}_1^* > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2)$ is still satisfied. By the Folk Theorem (Theorem 2.19), there is a $\delta \in (0, 1)$ and a subgame-perfect equilibrium (hence a Nash equilibrium) of $G_\delta^\infty$ with the normalized payoff profile $\tilde{\nu}^*$. This contradicts condition (2) in the definition of shallow game.

The following example shows that there are both shallow games where every computable strategy that is part of a subgame-perfect equilibrium has a best response, and shallow games where this is false.

Example 3.19. Let $G, H$ be games with $A_G = A_H = A = \{x, y\}^2$ and the payoff matrices below.

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>y</td>
<td>1, 0</td>
<td>3, 2</td>
</tr>
</tbody>
</table>

Table 2. Payoff matrix of $G$

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>0, 1</td>
<td>2, 2</td>
</tr>
<tr>
<td>y</td>
<td>1, 0</td>
<td>3, 2</td>
</tr>
</tbody>
</table>

Table 3. Payoff matrix of $H$
Observe that \((1, 2)\) is the minmax payoff profile in both games and that the action profile \((y, y)\) is a Nash equilibrium of both games, whence repeated play of \((y, y)\) is a subgame-perfect equilibrium in both \(G^\infty_\delta\) and \(H^\infty_\delta\) for any \(\delta\). Let \(\textbf{s}\) be the strategy profile where both players play \((y, y)\) forever, and note that, for any \(\delta\), we have \(v_1(\textbf{s}) = 3/(1-\delta) > \min_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2)/(1-\delta)\). Furthermore, for any \(\delta\) and any strategy profile \(\textbf{s}'\) in \(G^\infty_\delta\), as the maximum payoff for \(P_2\) is 2 in the stage game, we have \(v_2(\textbf{s}') \leq 2/(1-\delta) = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)/(1-\delta)\), and thus if \(\textbf{s}'\) is a Nash equilibrium, \(v_2(\textbf{s}') = 2/(1-\delta) = \min_{a_1 \in A_1, a_2 \in A_2} \max_{a_1 \in A_1} \min_{a_2 \in A_2} u_2(a_1, a_2)/(1-\delta)\). Hence, both \(G\) and \(H\) are shallow.

Under a subgame-perfect equilibrium of \(H^\infty_\delta\), \(P_2\) cannot threaten \(P_1\) with playing \(x\) because this will decrease the payoff of \(P_2\) (regardless of what \(P_1\) plays), whence the threat would not be credible. Thus, \(P_2\) always plays \(y\), and playing \(y\) forever is a best response for \(P_1\) and is clearly a computable strategy. This is not the case of \(G^\infty_\delta\) where \(P_2\) could threaten Player 1 to play \(x\) for a finite number of stages if they deviate.

The transformation defined below effectively removes the columns from the payoff matrix of \(G\) that contain only action profiles where the payoff of \(P_2\) is strictly smaller than their minmax payoff.

**Definition 3.20.** Let \(G = (\{1, 2\}, A, u)\) be a 2-player normal-form game. Define the canonical reduction of \(G\) to be the 2-player normal-form game \(G^{\min} = (\{1, 2\}, A^{\min}, u^{\min})\) with:

1. \(A^{\min}_1 = A_1\)
2. \(A^{\min}_2 = \{a' \in A_2 : \max_{a_1 \in A_1} u_2(a_1, a'_2) \geq \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)\}\)
3. \(u^{\min}(a) = u(a)\), for all \(a \in A^{\min}\)

For illustration, let us apply the canonical reduction to the games from Example 3.19. For the game \(G\), simply \(G^{\min} = G\). In \(H^{\min}\), however, \(P_2\) can only play \(y\), and the payoff matrix of \(H^{\min}\) has only one column:

<table>
<thead>
<tr>
<th>(x)</th>
<th>(y)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 2</td>
<td>3, 2</td>
</tr>
</tbody>
</table>

Table 4. Payoff matrix of \(H^{\min}\)

Although \(H\) is shallow, the minmax payoff profile in \(H^{\min}\) is \((3, 2)\), hence \(P_1\) cannot earn more than their minmax payoff and \(H^{\min}\) is not shallow. However, for every shallow game \(G\) it follows immediately from the definition of \(G^{\min}\) that there is the following useful relationship between the minmax payoffs of \(G\) and \(G^{\min}\):

**Observation 2.** Let \(G = (\{1, 2\}, A, u)\) be shallow. Then:

1. \(\min_{a_1 \in A_1^{\min}} \max_{a_2 \in A_2^{\min}} u^1_{\min}(a_1, a_2) \geq \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2)\)
2. \(\min_{a_1 \in A_1^{\min}} \max_{a_2 \in A_2^{\min}} u^2_{\min}(a_1, a_2) = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)\)

We next prove that if a shallow game \(G\) has a computable strategy for its infinitely repeated version that has no computable best response, then its canonical reduction \(G^{\min}\) must be shallow.

**Lemma 3.21.** Let \(G\) be a shallow game and \(G^{\min}\) its canonical reduction. If there is \(\delta\) with \(0 < \delta < 1\) such that for any \(\delta'\) with \(\delta < \delta' < 1\) there is a strategy profile \(s = (s_1, s_2)\) of \(G^\infty_\delta\) satisfying...
(1) s is a subgame-perfect equilibrium of $G^\infty_\delta$.
(2) $s_2$ is computable.
(3) $s_2$ does not have a computable best response,
then $G^{min}$ is shallow.

Proof. As $G$ is shallow, by Proposition 3.18, every $a \in A$ satisfies $u_2(a) \leq \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)$, and as $s$ is a subgame-perfect equilibrium of $G^\infty_\delta$, we have $s_2(h^T) \in A^{min}_2$ for all $h^T \in \mathcal{H}^T_{G^\infty_\delta}$. Therefore, $s$ is also a subgame-perfect equilibrium of $(G^{min})^\infty_\delta$. Hence, if $s_2$ has a best response $s_2'$ in $(G^{min})^\infty_\delta$, $v_1(s_2') = v_1(s_1)$. Moreover, $s_2$ is computable, and because any strategy in $(G^{min})^\infty_\delta$ is also a strategy in $G^\infty_\delta$, no computable best response to $s_2$ exists in $(G^{min})^\infty_\delta$ as it would also be a computable best response in $G^\infty_\delta$, contradicting the assumptions. Applying Lemma 3.1 to $G^{min}$ we obtain a Nash equilibrium $s'$ of $(G^{min})^\infty_\delta$ such that:

$$v_1(s') > \frac{1}{1 - \delta} \min_{a_1 \in A^{min}_1} \max_{a_2 \in A^{min}_2} u_1^{min}(a_1, a_2)$$

Because $G$ is shallow and by Observation 2, every Nash equilibrium $s''$ of $(G^{min})^\infty_\delta$ satisfies

$$v_2(s'') = \frac{1}{1 - \delta} \min_{a_1 \in A^{min}_1} \max_{a_2 \in A^{min}_2} u_2^{min}(a_1, a_2)$$

Thus, $G^{min}$ is shallow.

As the final lemma in this section, we prove that if the canonical reduction of a shallow game is shallow, then there is a computable strategy having no computable best response.

Lemma 3.22. Let $G$ be a shallow game such that $G^{min}$ is shallow. There is a $\delta$ with $0 < \delta < 1$ such that for any $\delta$ with $\delta < \delta < 1$, there is a strategy profile $s = (s_1, s_2)$ of $G^\infty_\delta$ such that:

(1) s is a subgame-perfect equilibrium of $G^\infty_\delta$.
(2) $s_2$ is computable.
(3) $s_2$ does not have a computable best response.

Proof. First observe that every subgame-perfect equilibrium of $(G^{min})^\infty_\delta$ is also a subgame-perfect equilibrium of $G^\infty_\delta$ (because $P_1$ can take the same actions in $G$ and $G^{min}$, and any action played by $P_2$ outside $G^{min}$ can only decrease their payoff below their minmax payoff).

It thus suffices to find a strategy profile in $(G^{min})^\infty_\delta$ that satisfies all three conditions of the lemma.

Let $m_1 = \min_{a_2 \in A^{min}_2} \max_{a_1 \in A^{min}_1} u_1^{min}(a_1, a_2)$ and $m_2 = \min_{a_1 \in A^{min}_1} \max_{a_2 \in A^{min}_2} u_2^{min}(a_1, a_2)$. Define $x \in A^{min}$ to be an action profile satisfying $u_1^{min}(x) = m_2$ and maximizing $u_1^{min}(x)$. Because $G^{min}$ is shallow, we have $u_1^{min}(x) > m_1$. Let $y_2 \in A^{min}_2$ be an action with $\max_{a_1 \in A^{min}_1} u_1^{min}(a_1, y_2) = m_1$ and let $y_1 \in A^{min}_1$ be an action with $u_2^{min}(y_1, y_2) = \max_{a_1 \in A^{min}_1} u_2^{min}(a_1, y_2)$. Observe that $\max_{a_1 \in A^{min}_1} u_2^{min}(a_1, y_2) \geq m_2$ because $G^{min}$ is a canonical reduction and $\max_{a_1 \in A^{min}_1} u_2^{min}(a_1, y_2) \leq m_2$ by Proposition 3.18, hence $u_2^{min}(y_1, y_2) = m_2$. Define $y = (y_1, y_2)$, so $u_1^{min}(y) \leq m_1$ and $u_2^{min}(y) = m_2$.

For any given $T \in \mathbb{N}$, let us define finite histories $h^T_C, h^T_D, h^T_E, h^T_F \in \mathcal{H}^T_{(G^{min})^\infty_\delta}$, as:

- $h^T_C[t] = x$ for each $t \in \{1, \ldots, T\}$
- $h^T_D[1] = y$ and $h^T_D[t] = x$ for each $t \in \{2, \ldots, T\}$
- $h^T_E[1] = h^T_E[2] = y$ and $h^T_E[t] = x$ for each $t \in \{3, \ldots, T\}$
• \( h^T_E[t] = y \) for each \( t \in \{3, \ldots, T\} \)

Define \( s^p \) to be the following strategy profile:

1. If no player has deviated in the last \( T \) stages, both players repeat the sequence \( h^T_E \).
2. If P1 has deviated at least once in the last \( T \) stages, both players play the punishment action \( y \).
3. Deviations by P2 are ignored.

Thus, \( s^p \) plays \( y \) for \( T \) stages from the last deviation by P1 and then returns to playing the sequence \( h^T_E \) forever. Choose \( T' \) and \( \delta' < 1 \) large enough that

\[
v_1(h^T_E) - v_1(h^T_E) > \eta_{G^{\min}}.
\]

This ensures that any one-shot deviation from \( s^p \) by Player 1, following any finite history, is unprofitable to Player 1. Because \( G^{\min} \) is shallow and P2 earns their minmax by playing \( s^p \), the one-shot deviation principle (Theorem 2.17) applies, and thus \( s^p \) is a subgame-perfect equilibrium.

Hence, we have obtained \( T' \), a \( \delta' \) with \( 0 < \delta < 1 \), finite histories \( h^T_C, h^T_D \), and a strategy profile \( s^p \) such that the following three conditions are satisfied for any \( \delta \) with \( \delta' < \delta < 1 \):

1. \( v_1(h^T_C) = \sum_{t=1}^{T} \delta'^{-1} u_1(h^T_C[t]) > \sum_{t=1}^{T} \delta^{-1} u_1(h^T_D[t]) = v_1(h^T_D) \)
2. \( s^p \) is a subgame-perfect equilibrium of \( (G^{\min})_{\delta} \) for any \( \delta \) with \( \delta' < \delta < 1 \); strategies \( s_1^p, s_2^p \) are both computable, and the continuation payoffs satisfy:

\[
v_1(s^p; h) < \frac{1}{1 - \delta'} v_1(h^T_D) \forall h \in H_{(G^{\min})_{\delta}},
\]
3. For each \( h \in H_{(G^{\min})_{\delta}} \), we have:

\[
v_2(h^T_C) = v_2(h^T_D) = (1 - \delta') v_2(s^p; h)
\]

The rest of the proof is the same as in the case of Theorem 3.4. Observe that strategies \( \sigma_1 \) and \( \sigma_2 \) with above-defined length of a block \( T \) and sequences of actions \( C_1, C_2, D_1, D_2 \) defined using the redefined histories \( h^T_C \) and \( h^T_D \) satisfy the conditions of Lemma 3.13. Moreover, by the observations above, \( \sigma_2 \) is computable and has no computable best response. Finally, Lemma 3.14 implies that \( \sigma_1 \) is a best response to \( \sigma_2 \) in every subgame, and because \( G^{\min} \) is shallow, \( \sigma_2 \) is a best response to \( \sigma_1 \) in every subgame. Hence \( (\sigma_1, \sigma_2) \) is a subgame-perfect equilibrium of \( (G^{\min})_{\delta} \).

**3.4 A complete characterization**

Putting it all together, we obtain a complete characterization of games having a computable strategy with no computable best response with conditions involving the existence of certain Nash equilibria of the repeated game, parallel to the characterization obtained for infinitely repeated games with the limit-of-means payoff [10]. This is used in the next section to derive the main result involving only payoffs of the stage game.

**Theorem 3.23.** Let \( G \) be a 2-player normal-form game, and let \( G^{\min} = (\{1, 2\}, A^{\min}, u^{\min}) \) be its canonical reduction. The following conditions are equivalent:

(a) There is a \( \delta \) with \( 0 < \delta < 1 \) such that for all \( \delta \) with \( \delta' < \delta < 1 \) there is a strategy profile \( s = (s_1, s_2) \) such that the following hold:

1. \( s \) is a subgame-perfect equilibrium of \( G_{\delta}^{\infty} \)
2. \( s_2 \) is computable,
3. \( s_2 \) does not have a computable best response.
Theorem 3.24. Let $G$ be a 2-player normal-form game, and let $G^{\text{min}} = (\{1, 2\}, A^{\text{min}}, u^{\text{min}})$ be its canonical reduction. The following conditions are equivalent:

(a) There is a $\delta$ with $0 < \delta < 1$ such that for all $\delta$ with $\bar{\delta} < \delta < 1$ there is a strategy profile $s = (s_1, s_2)$ such that the following hold:

1. $s$ is a subgame-perfect equilibrium of $G^\infty$.
2. $s_2$ is computable,
3. $s_2$ does not have a computable best response.

(b) There is a feasible and individually rational payoff profile $v'$ in $G$ that satisfies:

$$v'_1 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2),$$

and at least one of the following two conditions:

$$v'_2 > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2),$$

$$v'_1 > \min_{a_2 \in A_2^{\text{min}}} \max_{a_1 \in A_1^{\text{min}}} u_1^{\text{min}}(a_1, a_2).$$

Proof. (a) $\Rightarrow$ (b): By Theorem 3.23, either:

1. There is a $\delta'$ and a Nash equilibrium $s'$ of $G^\infty_0$ that satisfies:

$$v_1(s') > \frac{1}{1 - \delta'} \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2),$$

and at least one of the following two conditions:

$$v_2(s') > \frac{1}{1 - \delta'} \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)$$

$$v_1(s') > \frac{1}{1 - \delta'} \min_{a_1 \in A_1^{\text{min}}} \max_{a_2 \in A_2} u_1^{\text{min}}(a_1, a_2).$$

(b) $\Rightarrow$ (a): If both (9) and (10) hold, Theorem 3.4 furnishes existence of an $s$ satisfying the conditions of (a). If there is no Nash equilibrium satisfying both (9) and (10), but there is a Nash equilibrium satisfying both (9) and (11), then both $G$ and $G^{\text{min}}$ are shallow, whence (a) holds by Lemma 3.22.

3.5 The main result

We can now finally prove the main result of the paper:

(2) There is a δ′ and a Nash equilibrium s′ satisfying the inequalities (9) and (11), that is, the normalized payoff profile ̄ν(s′) satisfies the inequalities (12) and (14). Because s′ is a Nash equilibrium, by The Folk Theorem, ̄ν(s′) is feasible.

In any case, ̄ν(s′) satisfies the condition (12) and one of (13) and (14), whence (b) holds.

(b) ⇒ (a): Let us consider the following two distinct cases:

(1) If ν′ satisfies both (12) and (13), ν′ is strictly individually rational. By The Folk Theorem, there is a δ′ and a Nash equilibrium s′ satisfying ̄ν(s′) = ν′, whence such s′ and δ′ satisfy the inequalities (9) and (10). Applying Theorem 3.23, (a) holds.

(2) There is a feasible and individually rational payoff profile ν′ satisfying the inequalities (12) and (14), but no strictly individually rational payoff profile is feasible in G. Observe that every action profile a ∈ A satisfies u_2(a) ≤ min \max_{a_1 \in A_1, a_2 \in A_2} u_2(a_1, a_2), as otherwise, for a′ with u_2(a′) > min \max_{a_1 \in A_1, a_2 \in A_2} u_2(a_1, a_2) and a small enough ε ∈ (0, 1), the payoff profile (1 − ε)ν′ + εν′(a′) is feasible and strictly individually rational. Define x ∈ A to be an action profile satisfying u_2(x) = min \max_{a_1 \in A_1, a_2 \in A_2} u_2(a_1, a_2) and maximizing u_1(x). Because ν′ is a convex combination of payoff profiles a with u_2(a) = u_2(x), u_1(x) ≥ ν′ > min \max_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2). Define y_2 ∈ A_2 to be the minmax action against P1, that is max u_1(a_1, y_2) = min \max_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2).

Consider the strategy profile s′ where P1 always plays x_1 and P2 plays the following trigger strategy – play x_2 if P1 has played x_1 in all previous rounds, otherwise play y_2 forever. This is a standard construction of trigger strategies – for large enough discount factor δ′, no player has an incentive to deviate, hence s′ is a Nash equilibrium of G^∞_δ with u_1(s′) > \frac{1}{1-\delta} min \max_{a_1 \in A_1, a_2 \in A_2} u_1(a_1, a_2). At the same time, for every δ, every Nash equilibrium s″ of G^∞_δ satisfies u_2(s″) = \frac{1}{1-\delta} min \max_{a_1 \in A_1, a_2 \in A_2} u_2(a_1, a_2), hence G is shallow by definition. Using the inequality (14) and Observation 2, we can analogously prove that G^min is shallow. By Lemma 3.22, (a) holds.

\[\square\]

3.6 Some examples

Below we give examples of games (and their canonical reductions) satisfying subsets of the conditions in (b). The examples have been chosen to show that all conditions in Theorem 3.24 are necessary, and that games satisfying different combinations of these conditions exist. Moreover, Examples 3.25 and 3.27 illustrate games that do not satisfy the sufficient conditions stated in Nachbar and Zame’s work [22].

Example 3.25. Let G_1 be a two-player game with A_1 = A_2 = \{a, b\} and the payoff matrix depicted in Table 5. The minmax payoff profile of G_1 is (1, 2) and 2 is the maximum payoff that P2 can obtain at any stage, so the inequality (13) is not satisfied. The feasible payoff profile (3, 2) is strictly individually rational for P1, satisfying the condition (12). In the canonical reduction G^min_1, \min_{a_2 \in A_2^2} \max_{a_1 \in A_1^1} u_1(a_1, a_2) = 1 < 3, and the condition (14) of Theorem 3.24 is also satisfied. Consequently, (a) holds.

Example 3.26. Let G_2 be a two-player game with A_1 = A_2 = \{a, b\} and the payoff matrix depicted in Table 6. The minmax payoff profile of G_2 is (1, 2). Condition (13) is not satisfied because P2 cannot obtain more than their minmax payoff. The minmax payoff profile of the canonical reduction G^min_2 is (3, 2), so also the condition (14) is not satisfied, as 3 is the maximum payoff that P1 can obtain. By Theorem 3.24, (a) does not hold.
Discounted Repeated Games having Computable Strategies with no Computable Best Response under Subgame-Perfect Equilibria

Table 5. Payoff matrix of $G_1$ and its canonical reduction, $G_1^\text{min}$.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0,2</td>
<td>2,2</td>
</tr>
<tr>
<td>b</td>
<td>1,0</td>
<td>3,2</td>
</tr>
</tbody>
</table>

Table 6. Payoff matrix of $G_2$ and its canonical reduction, $G_2^\text{min}$.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0,1</td>
<td>2,2</td>
</tr>
<tr>
<td>b</td>
<td>1,0</td>
<td>3,2</td>
</tr>
</tbody>
</table>

Example 3.27. Let $G_3$ be a two-player game with $A_1 = A_2 = \{ a, b \}$ and the payoff matrix depicted in Table 7. The minmax payoff profile of $G_3$, as well as of $G_3^\text{min}$ is $(2, 1)$. We can see that the condition (14) is not satisfied as 2 is the maximum payoff that $P_1$ can obtain in $G_3^\text{min}$. For example, the payoff profile $(2.2, 1.6) > (2, 1)$ is feasible and strictly individually rational. Hence, $s$ satisfies the conditions (12) and (13), and by Theorem 3.24, (a) holds.

Table 7. Payoff matrix of $G_3$ and its canonical reduction, $G_3^\text{min}$.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>3,0</td>
<td>2,2</td>
</tr>
<tr>
<td>b</td>
<td>3,0</td>
<td>2,1</td>
</tr>
</tbody>
</table>

3.7 The search for a best response

So far, we studied the complexity of strategies and their best responses, with the assumption that the players already have the strategies at hand. We now show that for the same class of games as in Theorem 3.24, the problem of finding a best computable response to a computable strategy is not computable.

Definition 3.28. Let $G$ be a 2-player normal-form game and $\delta \in (0, 1)$. Define $C_\delta$ to be the set of computable strategies of $P_2$ in $G_\delta^\infty$ that admit a computable best response.

Definition 3.29. Let $G$ be a 2-player normal-form game and $\delta \in (0, 1)$. Define a function $\text{BR} : C_\delta \rightarrow (\mathcal{H} \rightarrow A_1)$ that for a given $s_2 \in C_\delta$, returns $\text{BR}(s_2)$, a best response to $s_2$. We say that $\text{BR}$ is computable if there is a Turing machine $M$ computing $\text{BR}$, that is, $M$ will on input (an encoding of) a Turing machine computing $s_2$, output (an encoding of) a Turing machine computing a best response to $s_2$.

Theorem 3.30. Let $G$ be a 2-player normal-form game, and let $G^\text{min} = (\{ 1, 2 \}, A^\text{min}, u^\text{min})$ be its canonical reduction. Assume that the conditions of (b) in Theorem 3.24 hold, that is, there exists a feasible and individually rational payoff profile $v'$ in $G$ that satisfies:

$$v'_1 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2),$$

and at least one of the following two conditions:

$$v'_2 > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)$$

Then there is a \( \delta \) with \( 0 < \delta < 1 \) such that for all \( \delta \) with \( \delta < \delta < 1 \), there is no computable function \( \text{BR} : C_\delta \to (H \to A_1) \).

Proof. We already proved that, for large enough \( \delta \), the computable strategy \( \sigma_2 \) from Definition 3.10 has no computable best response. Recall that for each \( i \in \mathbb{N} \), in the \( i \)-th test period, \( P_1 \) plays the sequence \( C_i \) if \( i \in A \) and \( D_i \) if \( i \in B \). Now, for each \( n \in \mathbb{N} \), define \( \sigma^n_2 \) to be the same strategy as \( \sigma_2 \), with one exception: in the \( i \)-th test period, the prescribed sequence of \( P_1 \) is \( C_i \) if \( n \in A \) and \( D_i \) if \( n \in B \). That is, for a fixed \( n \), the strategy \( \sigma^n_2 \) expects \( P_1 \) to play the same sequence of actions in each test period. Observe that for any \( n \in \mathbb{N} \), \( \sigma^n_2 \) has a computable best response \( \sigma^n_1 \) – the answers to \( n \in A \) and \( n \in A \) can be saved in the description of the Turing machine computing \( \sigma^n_1 \). This implies that for each \( n \), \( \sigma^n_2 \in C_\delta \). However, if there were a Turing machine \( TM_{\text{BR}} \) computing \( \text{BR} \), \( TM_{\text{BR}} \) would compute a best response to \( \sigma^n_2 \) for each \( n \in \mathbb{N} \), which involves computing whether \( n \in A \) or \( n \in B \). Therefore, we could use \( TM_{\text{BR}} \) to separate sets \( A \) and \( B \), contradicting Proposition 2.23. Hence, no computable \( \text{BR} \) exists.

Theorem 3.30 is a generalization of [22, Theorem 2], where a similar result is shown for the stronger assumptions of Nachbar and Zame.

Observe that a partial “converse” to Theorem 3.30 holds: if every feasible and individually rational payoff profile \( v \) satisfies \( v'_1 = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2) \) then the proof of Lemma 3.1 provides us with a simple algorithm to compute a best response to any computable strategy, and hence \( \text{BR} \) is computable. However, we do not expect in general that if the conditions of Theorem 3.30 fail to hold, it will always imply that \( \text{BR} \) is computable; we leave investigation of this possibility to future work.

4 POLYNOMIAL-TIME DECIDABILITY

We proceed to show that conditions of Theorem 3.24 are decidable in polynomial time in the size of the payoff matrix of \( G \).

THEOREM 4.1. It is decidable in polynomial time in \( |A_1| \times |A_2| \) whether a two-player normal-form game \( G = \{\{1, 2\}, A = A_1 \times A_2, u\} \) satisfies conditions (b) of Theorem 3.24.

Proof. For the concrete upper bounds in the proof, we assume that each payoff in the matrix is a rational number that can be represented in a machine word; if payoffs are rational, but cannot be represented in a single word, standard bitwise comparisons and arithmetic can be performed in polynomial time, and the result follows.

We claim that Algorithm 2, further below, decides whether the conditions of Theorem 3.24 are satisfied. The algorithm makes several calls to Algorithm 1, which for a given game \( G \) decides whether the set of feasible payoff profiles in \( G \) contains one that is individually rational for \( P_2 \) and strictly individually rational for \( P_1 \).

We now describe Algorithm 1 in more detail. A payoff profile \( v \in \mathbb{R}^{|A_1|} \) is feasible if it is in the convex hull \( C \) of the payoff profiles of \( G \), \( v_{a \in A}(u(a)) \). The set of individually rational payoff profiles is

\[
Q = \{(x, y) \in \mathbb{R}^2 : x = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_1(a_1, a_2), y = \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)\}
\]

Our goal is to decide if there is a payoff profile \( v \in C \cap Q \) satisfying \( v_1 > \min_{a_1 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2) \). To do this, we solve the linear program on Line 3 – from the set of all feasible and individually rational payoff profiles, if non-empty, the program outputs a payoff profile \( v' \) with \( v'_1 \) maximized.
Algorithm 1: HasStrictPayoff1

Input: Game $G = (\{1, 2\}, A, u)$.
Output: Boolean: there is a feasible and individually rational payoff profile $v$ with $v_1 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$.

1 $C = \text{ConvexHull}(\cup_{a \in A} \{u(a)\})$;
2 $Q = \{(x, y) \in \mathbb{R}^2 : x \geq \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2), y \geq \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)\}$;
3 $v_1^* = \max_{(x, y) \in C \cap Q} x$;
4 if the linear program is infeasible then
   return False;
5 if $v_1^* > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$ then
   return True;
6 else
   return False;
end

If $v_1^* > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$, then $v^*$ is the payoff profile we were searching. Otherwise, no such payoff profile exists – either the linear program is infeasible (equivalent to $C \cap Q = \emptyset$), or for all $v \in C \cap Q : v_1 = \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$.

The condition $(x, y) \in C$ can be expressed as a set of conditions $(x, y) = \sum_{a \in A} \alpha_a u(a)$, where $\forall a \in A : \alpha_a \leq 1$ and $\sum_{a \in A} \alpha_a = 1$. Hence, the linear program is bounded, contains $nm + 2$ variables and $nm + 5$ constraints. Using, e.g., Vaidya’s algorithm [30] for solving the linear program, we obtain the solution in time $O((nm)^{2.5})$. In total, Algorithm 1 computes whether there is a feasible and individually rational payoff profile $v$ with $v_1 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)$ in time $O((nm)^{2.5})$.

Algorithm 2: Verify

Input: Game $G = (\{1, 2\}, A, u)$.

1 if HasStrictPayoff1($G$) then
2   $G^t = \text{Transpose}(G)$;
3   if HasStrictPayoff1($G^t$) then
4     return True;
5 else
6   $G^{\text{min}} = \text{Reduce}(G)$;
7   return HasStrictPayoff1($G^{\text{min}}$);
8 end
9 else
10   return False;
11 end

Note that the running time of Vaidya’s algorithm is $O((nm)^{2.5} L)$ where $L$ is bounded by the number of bits in the input. With our assumption that each payoff can be represented in a machine word, this becomes $O((nm)^{2.5})$. 
Algorithm 2—the verification algorithm—decides whether a given game satisfies the condition of Theorem 3.24. Line 1 computes whether there is a feasible and individually rational payoff profile \(v\) with \(v_1 > \min_{a_2 \in A_2} \max_{a_1 \in A_1} u_1(a_1, a_2)\). If this is not satisfied, the condition (12) of Theorem 3.24 does not hold. If the condition in line 1 is satisfied, we compute \(G'\), the game \(G\) with roles of players interchanged, that is, with the payoff matrix transposed. Line 3 hence computes whether there is a feasible and individually rational payoff profile \(v'\) with \(v'_2 > \min_{a_1 \in A_1} \max_{a_2 \in A_2} u_2(a_1, a_2)\).

If there is such a payoff profile, then \(v^* = \frac{1}{2}v + \frac{1}{2}v'\) is feasible and strictly individually rational payoff profile, satisfying both conditions (12) and (13), hence the condition (b) holds.

Otherwise, if the condition in line 3 does not hold, but the condition in line 1 holds, the game \(G\) is shallow. Line 6 computes the canonical reduction \(G^{\min}\)—which can be done by first computing the values \(\max_{a_2 \in A_2} u_2(a_1, a_2)\) for each \(a_1 \in A_1\); then computing the minimum of these values, which equals P2’s minmax value; and finally removing columns with maximum P2’s payoff smaller than P2’s minmax value. Line 7 computes whether the condition (14) holds— if it does, (b) is satisfied, otherwise (b) is not.

For a game \(G\) with a payoff matrix of dimension \(n \times m\), the algorithm calls Algorithm 1 at most three times, which takes \(O\((nm)^{2.5}\) time. Both Transpose\((G)\) and Reduce\((G)\) can be easily computed in \(O(nm)\) time. Hence, Algorithm 2 computes whether \(G\) satisfies the conditions of Theorem 3.24 in \(O\((nm)^{2.5}\) time.

□

Acknowledgments

We are indebted to the anonymous referees for their diligent work that has helped improve the paper. In particular, one referee’s extensive and very careful review has resulted in substantial improvement of the presentation of our results.
A APPENDIX: PROOF SKETCHES OF SOME STANDARD AUXILIARY RESULTS

This appendix contains proof sketches of the particular version of the Folk Theorem and the One-Shot Deviation Principle that we use in the paper. In both cases, the proofs can be found elsewhere in the literature, but proof sketches are included here for completeness, with computability made explicit compared to the original proofs in [13, 20].

Proof Sketch of The Folk Theorem—Theorem 2.19. (1) The strategy of Pi that plays a best response to the opponent’s action in every stage, obtains payoff at least \( \min_{a_i \in A_i} \max_{a_{-i} \in A_{-i}} u_i(a_i, a_{-i}) \) in every stage, so each Pi earns at least \( \frac{1}{1-\delta} \min_{a_i \in A_i} \max_{a_{-i} \in A_{-i}} u_i(a_i, a_{-i}) \) in \( G^\infty \). Therefore, this is the minimum \( \delta \)-discounted payoff under a Nash equilibrium, and thus \( v \) is individually rational. Let \( s \) be a Nash equilibrium with the \( \delta \)-discounted payoff \( v(s) = \frac{1}{1-\delta} v \), and let \( h^\infty_s \) be the path of play obtained by playing \( s \). Then:

\[
\hat{v} = (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} u(h^\infty_s[t])
\]

For each \( t \), we have \( h^\infty_s[t] \in A \), and thus \( (1-\delta) \sum_{t=1}^{\infty} \delta^{t-1} u(h^\infty_s[t]) = \sum_{a \in A} \beta a u(a) \) for some \( \beta \in \mathbb{R}^{|A|} \). Moreover, for all \( t \) we have \( \beta_0 u(\hat{s}) = 1 \), and thus \( \sum_{a \in A} \beta a = 1 \), so \( \hat{v} \) is a convex combination of the stage game payoff profile, and thus \( \hat{v} \) is feasible.

(2) First, for a feasible and strictly individually rational payoff profile \( \hat{v} \), we define a path of play \( h^\infty \) with the normalized discounted payoff \( \hat{v}(h^\infty) = (1-\delta) v \). Since \( \hat{v} \) is feasible, we can write \( \hat{v} = \sum_{a \in A} \alpha_a u(a) \) with \( 0 \leq \alpha_a \leq 1 \) and \( \sum_{a \in A} \alpha_a = 1 \). Write \( A = \{ a^1, \ldots, a^k \} \). We now describe the algorithm that, on input \( t \in \mathbb{N} \), computes \( h^\infty[t] \). Set \( I^0(t) = 1 \) if \( h^\infty[t] = a^k \), and set \( I^0(t) = 0 \) otherwise; furthermore, let \( N^k(1) = 0 \) for all \( k = 1, \ldots, |A| \). First compute

\[
N^k(t) = \sum_{t'=1}^{t-1} (1-\delta) \delta^{t'-1} I^k(t'),
\]

then

\[
C(t) = \{ k : \alpha_{a^k} - N^k(t) > \delta^{t-1} (1-\delta) \},
\]

and

\[
k^* (t) = \arg \max_{k \in C(t)} \{ \alpha_{a^k} - N^k(t) \}.
\]

Finally, set \( h^\infty[t] = a^{k^*(t)} \). Observe that if we define \( s \) to play \( h^\infty[t] \) in the stage \( t \) following any history, \( v(s) = \hat{v} \), but the continuation payoffs might be far from \( \hat{v} \). To remedy this, compute feasible payoff vectors close to \( \hat{v} \), and play finite sequences of a fixed length instead of playing \( a^{k^*(t)} \) in stage \( t \). Details of the construction and the proof that the continuation payoffs are within \( \epsilon \) of \( v \) can be found in [20]. Observe that every step of the algorithm is computable, and hence \( s \) is computable. Let us define a strategy profile \( v^s \) with \( v(v^s) = v \).

(A) If no player has deviated, play according to \( s \).

(B) Following a single deviation by Pi, both players play the minmax against Pi for a fixed number of \( K_m \) stages. Hence, Player \(-i \) plays \( m_{-i} = \arg \min_{a_{-i}} \max_{a_i \in A_i} u_i(a_i, m_{-i}) \), and Pi plays

\[
\arg \max_{a_i \in A_i} u_i(a_i, m_{-i})
\]

for \( K_m \) stages.

(C) Play a low-payoff path \( p^{(i)} \), dependent on the deviating player and on the payoff profile obtained in the punishment phase (B).
We omit the exact definition of $K_m$, as well as the paths of play needed in (C). Note, however, that $\sigma^u$ is computable, as the strategies in (A) and (C) are computable due to previous discussion and (B) is simply playing minmax for a fixed number of stages.

**Proof sketch of the one-shot deviation principle (Theorem 2.17).** If a single deviation from $s_t(h^T)$ leads to a higher payoff for $P_1$, it is not a subgame-perfect equilibrium – simply because the original strategy is not a best response to the other player’s strategy in the subgame $(G^{\infty}, h^T)$.

To prove the other direction, assume for a contradiction that no one-shot deviation profitable for $P_1$ (without loss of generality) exists, but $s$ is not a subgame-perfect equilibrium, i.e. there is a history $h^t \in \mathcal{H}^t_{G^{\infty}}$ such that $s$ is not a Nash equilibrium in the subgame $(G^{\infty}, h^t)$. That is, following $h^t$, player $1$ has a better strategy $s'_1$.

Let $r_1[1], r_1[2], \ldots$ denote the infinite sequence of payoffs for $P_1$ induced by $s$, and similarly, $r'_1[1], r'_1[2], \ldots$ induced by the strategy profile $s' = (s'_1, s_{-1})$. We assume that $s'_1$ gives a higher reward to $P_1$ in the subgame $(G^{\infty}, h^t)$ than $s_1$, i.e.

$$v_1(s'_1, s_{-1}; h^t) > v_1(s_1, s_{-1}; h^t),$$

so for some $\epsilon > 0$,

$$v_1(s'_1, s_{-1}; h^t) \geq v_1(s_1, s_{-1}; h^t) + 2\epsilon.$$

For a fixed $n \in \mathbb{N}$, let $s_1^n$ be a combined strategy in $(G^{\infty}, h^t)$ that tells $P_1$ to play according to $s'_1$ in stages $t + 1, \ldots, t + n$, and switch back to $s_1$ from stage $t + n + 1$. Again, let $r_1^n[t + 1], r_1^n[t + 2], \ldots$ be the payoffs induced by $s_1^n = (s_1^n, s_{-1})$. There exists $n_0 \in \mathbb{N}$ such that for any $n > n_0$ the payoffs in later stages become unimportant:

$$\left| \sum_{i=t+n+1}^{\infty} \delta^{i-1} r_1^n[i] - \sum_{i=t+n+1}^{\infty} \delta^{i-1} r_1[i] \right| < \epsilon,$$

hence

$$\sum_{i=t+n+1}^{\infty} \delta^{i-1} r_1^n[i] \geq \sum_{i=t+n+1}^{\infty} \delta^{i-1} r_1[i] + \epsilon,$$

and finally

$$v_1(s_1^n, s_{-1}; h^t) > v_1(s_1, s_{-1}; h^t). \quad (15)$$

That is, there is a strategy $s_1^n$ that is a better response to $s_{-1}$ than $s_1$ in $(G^{\infty}, h^t)$, and differs from $s_1$ only in a finite number of stages. Let $s_1^N$ be any strategy satisfying:

1. $v_1(s_1^N, s_{-1}; h^t) > v_1(s_1, s_{-1}; h^t)$,
2. $s_1^N$ copies $s_1$ from stage $t + N + 1$ and
3. for any $0 \leq N' < N$, any strategy $s_1^{N'}$ that copies $s_1$ from stage $t + N' + 1$ does not improve the payoff in $(G^{\infty}, h^t)$, i.e. $v_1(s_1^{N'}, s_{-1}; h^t) \leq v_1(s_1, s_{-1}; h^t)$.

Such a strategy clearly exists – the first condition is satisfied by $s_1^n$, the set of strategies satisfying the second condition is finite (and includes $s_1^n$), and $N = 1$ satisfies the last condition, since for $N' = 0$, $s_1^N$ coincides with $s_1$.

Let $h^{t+N-1}$ be the $(t+N-1)$-period history achieved by playing $(s_1^N, s_{-1})$ in the subgame $(G^{\infty}, h^t)$. Because $s_1^N$ is a one-shot deviation from $s_1$ in $(G^{\infty}, h^{t+N-1})$ (by condition (2)), it does not strictly increase the payoff of $P_1$:

$$v_1(s_1, s_{-1}; h^{t+N-1}) \geq v_1(s_1^N, s_{-1}; h^{t+N-1}). \quad (16)$$
If \( N = 1 \), we can modify \( s_N^1 \) to be the strategy that copies \( s_1 \), except that we keep \( s_N^1(h^t) \), hence \( s_1^1(h^t) \neq s_1(h^t) \) and still \( v_1(s_N^1, s_{-1}; h^t) > v_1(s_1, s_{-1}; h^t) \). However, such \( s_N^1 \) is a one-shot deviation from \( s_1 \), which contradicts our assumption that no one-shot deviation profitable for player one exists.

Otherwise, if \( N > 1 \), let \( s_N^{N-1} \) be a strategy that copies \( s_N^1 \) but plays according to \( s_1 \) in stage \( (t + N) \). Playing \( (s_N^{N-1}, s_{-1}) \) in \( (G^\infty, h^t) \) also achieves the history \( h^{t+N-1} \), so by condition (1) above and inequality (16) we obtain

\[
v_1(s_N^{N-1}, s_{-1}; h^t) \geq v_1(s_N^1, s_{-1}; h^t) > v_1(s_1, s_{-1}; h^t).
\]

This contradicts condition (3) for strategy \( s_N^1 \).
REFERENCES


