Les Cahiers du GERAD ISSN: 0711–2440

Approximate perfect equilibria in finitely repeated Prisoner's Dilemma with asymmetric players

E. Parilina, A. Pisareva, G. Zaccour

G–2026–21

April 2026

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2026
– Bibliothèque et Archives Canada, 2026

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2026
– Library and Archives Canada, 2026

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Approximate perfect equilibria in finitely repeated Prisoner's Dilemma with asymmetric players

Elena Parilina [a]

Alena Pisareva [a]

Georges Zaccour [b, c]

[a] *Saint Petersburg State University, Saint Petersburg, Russia*

[b] *Chair in Game Theory and Management & GERAD, Montréal (Qc), Canada, H3T 1J4*

[c] *Department of Decision Sciences, HEC Montréal, Montréal (Qc), Canada, H3T 2A7*

e.parilina@spbu.ru
st055836@student.spbu.ru
georges.zaccour@gerad.ca

April 2026
Les Cahiers du GERAD
G–2026–21

Abstract : In this paper, we study a finitely repeated Prisoner's Dilemma in which players are asymmetric in both their temptation to deviate from cooperation and their level of patience, as captured by their discount factors. We investigate whether a profile of limited retaliation strategies constitutes a perfect ε-equilibrium. These strategies impose a mild punishment on the first player to deviate and allow cooperation to resume after a finite retaliation period. Any subsequent deviation is punished until the end of the game. A key feature of these strategies is that, when it exists, the duration of the retaliation period need not be unique. We characterize both ex ante and contemporaneous perfect ε-equilibria and show how the duration of the retaliation period depends on asymmetries in the game's parameters.

Keywords : Prisoner's Dilemma; repeated games; asymmetric players; approximate equilibrium

Résumé : Dans cet article, nous étudions un dilemme du prisonnier répété un nombre fini de fois, où les joueurs présentent une asymétrie tant dans leur propension à dévier de la coopération que dans leur niveau de patience, ce dernier étant reflété par leurs facteurs d'actualisation. Nous examinons si un ensemble de stratégies de représailles limitées constitue un ε-équilibre parfait. Ces stratégies imposent une sanction modérée au premier joueur qui dévie et permettent la reprise de la coopération après une période de représailles finie. Toute déviation ultérieure est sanctionnée jusqu'à la fin du jeu. Une caractéristique essentielle de ces stratégies est que, lorsqu'elle existe, la durée de la période de représailles n'est pas nécessairement unique. Nous caractérisons les ε-équilibres parfaits ex ante et contemporains et montrons comment la durée de la période de représailles dépend des asymétries des paramètres du jeu.

Mots clés : Dilemme du Prisonnier; jeux répétés; joueurs asymétriques; équilibre approximatif

Acknowledgements: The work of the first and second authors is supported by the Russian Science Foundation, grant no. 25–21–00581. The work on characterization of the subgame perfect ε-equilibrium and numerical analysis is supported by the Russian Science Foundation, grant no. 25–21–00581. The third author acknowledges the financial support by NSERC, Canada, Grant RGPIN–2021–02462.

1 Introduction

The strategic interaction among rational agents, where current decisions take into account both past actions and future consequences, is commonly formalized within the theory of repeated games. This framework provides a powerful tool for analyzing the sustainability of cooperation, the emergence of reputation, and the efficiency of long-term contracts under strategic uncertainty [19].

The folk theorem establishes that, in infinitely repeated games with a sufficiently high discount factor, a wide range of outcomes, including cooperative ones, can be sustained as equilibrium behavior. However, these results do not carry over to games with a finite and known horizon. In particular, a standard backward-induction argument shows that cooperation cannot be sustained as part of a subgame-perfect Nash equilibrium [1, 11, 12]. Traditional enforcement mechanisms, such as grim trigger strategies, which prescribe permanent punishment following any deviation, have been criticized for their economic inefficiency and lack of robustness [1, 4]. In finite-horizon or incomplete-information repeated games, such strategies may lead to excessively harsh outcomes and can be fragile in the presence of random errors or asymmetries in players' preferences. In this context, it is natural to investigate strategies with limited-duration punishment schemes that allow cooperation to resume after deviations [22, 23], or to consider approximate equilibria, which provide a theoretically and practically appealing second-best solution. In an approximate equilibrium, no player can gain more than a predetermined bound from deviating from the prescribed strategy [25].

In this paper, we study a class of limited retaliation strategies (LRSs) in a finitely repeated Prisoner's Dilemma, allowing for asymmetries in stage payoffs, discount factors, or both. The duration of the punishment phase is determined endogenously as a function of the stage at which a deviation occurs, the length of the game, and the players' payoff structure. Compared to grim trigger strategies, LRSs are more flexible and arguably more realistic, particularly in finite-horizon settings. The main theoretical contribution of this work is to expand the toolkit for analyzing approximate subgame-perfect equilibria in finite dynamic games with asymmetric players.

Related ideas involving relaxations of grim trigger strategies are explored in [3], albeit in the context of infinitely repeated games. In that work, subgame-perfect equilibria with justified punishments are constructed. A punishment is considered justified if either the deviation harms the other player or the continuation payoff benefits the other player. The authors characterize the set of payoff vectors that can be supported by such equilibria as the discount factor approaches one.

Finally, we also present results for the undiscounted finitely repeated Prisoner's Dilemma. The absence of discounting is of particular interest because future payoffs are valued equally with present ones, which can significantly affect strategic behavior [15]. In such settings, players may be more inclined to cooperate in earlier stages, as the value of future rewards remains constant throughout the game. However, as the final stage approaches, the incentive to defect becomes stronger, making the analysis of these environments especially relevant for understanding the dynamics of cooperation and conflict.

1.1 Brief literature review

The literature on repeated games is vast, and an exhaustive review lies beyond the scope of this paper. Instead, we focus on asymmetric repeated games, and in particular on finitely repeated Prisoner's Dilemma (PD) games, which constitute the core of our analysis.

Asymmetries in PD games can arise in several ways. First, players may differ in their available strategy sets or in the externalities generated by their actions. Such asymmetries may stem, for example, from differences in players' positions within a network when they are randomly matched to play a PD at each stage, as commonly studied in evolutionary game theory; see, e.g., [9, 14].

Second, players may possess asymmetric information. Two-player repeated zero-sum games with asymmetric information, initially introduced in [26], are further analyzed in [6]. In that setting, the

state evolves according to a Markov process, and players have unequal access to information: one player observes the state before choosing an action, whereas the other never observes it. In [20], the authors study repeated games in which each player has incomplete information about the other's discount factor and derive necessary and sufficient conditions for sustaining full cooperation in equilibrium using grim trigger strategies.

Third, asymmetry may arise from differences in discount factors. The role of long- and short-lived players in shaping equilibrium outcomes is examined in [19]. Two-player repeated games with heterogeneous discounting are studied in [16], while [17] extends the analysis to settings with incomplete information, where one player is uncertain about the other's discount factor. In this case, asymmetry arises both from information and time preferences. The authors characterize equilibrium outcomes through the feasible payoff set under heterogeneous discounting. A folk theorem for infinitely repeated games with unequal discount factors is established in [7], showing that cooperation may still be sustained despite differences in patience. Similarly, [8] characterizes subgame-perfect equilibria in infinitely repeated PD games with heterogeneous discount factors. Asymmetry in discounting or payoffs typically leads to asymmetric punishment durations under limited punishment schemes, a feature consistent with our findings. In addition, [10] studies infinitely repeated PD games with side payments and characterize the Pareto frontier of subgame-perfect equilibrium payoffs for all possible combinations of discount factors.

Fourth, players may differ in their stage payoffs, even when they share the same strategy set, an asymmetry that can be interpreted as differences in "power." Relatively few papers address payoff asymmetry in repeated games. Some experimental studies examine its impact on cooperation and collusion. For instance, [2] investigates payoff asymmetry in laboratory PD games and finds "*asymmetry reduces the rates of cooperation in simultaneous games. In sequential games, asymmetry interacts with order of play such that the rate of cooperation is highest when payoff disadvantaged players move first.*" Extending this line of experimental research, [5] shows that in finitely repeated games, payoff asymmetry not only reduces overall cooperation but also introduces fairness considerations. Low-payoff players may defect to mitigate perceived inequity, while high-payoff players may tolerate such deviations. These findings highlight that payoff asymmetry introduces a behavioral dimension (inequality aversion) into strategic interactions.

Overall, asymmetry in repeated games has been studied less systematically than in symmetric settings (see, e.g., [19]). Differences in discount factors or payoff structures can significantly affect incentives for cooperation, the credibility of punishment strategies, and the nature of equilibrium paths. In particular, when one player is more impatient, the standard Folk Theorem conditions may fail or require longer punishment phases to sustain cooperation. Our objective is to analyze how such asymmetries influence the existence and stability of equilibrium outcomes, and how the required duration of punishment varies with the degree of heterogeneity between players.

Our analysis extends the framework developed in [21], where approximate equilibria were established for symmetric players under finite punishment schemes. Here, we investigate how relaxing the symmetry assumption modifies the sufficient conditions for sustaining cooperation in finitely repeated interactions.

The remainder of the paper is organized as follows. Section 2 introduces the model of a finitely repeated Prisoner's Dilemma with asymmetric players and defines the class of limited retaliation strategies, along with the corresponding punishment durations satisfying desired properties. Section 3 presents the main theoretical results, characterizing ex-ante and contemporaneous perfect ε-equilibria. Section 4 provides numerical examples illustrating how asymmetry affects the duration of punishment phases. Section 5 concludes.

2 Model

We consider a two-player Prisoner's Dilemma (PD) game repeated T times with the stage game payoffs:

	C	N
C	(a, a)	(c, b_2)
N	(b_1, c)	(d, d)

(1)

where the actions C and N stand for cooperate and do not cooperate, respectively. The payoffs satisfy the inequalities $b_i > a > d > c$ and $2a > b_i + c$ for $i \in I = \{1, 2\}$, and are interpreted as follows:

a : payoff of player $i = 1, 2$ when both players cooperate;

d : payoff of player $i = 1, 2$ when both players defect;

c : payoff of player i when she cooperates while player $3 - i$ does not, $i = 1, 2$;

b_i : payoff of player i when she defects while player $3 - i$ cooperates, $i = 1, 2$.

For both players, strategy N strictly dominates strategy C in the one-stage Prisoner's Dilemma game, so the only Nash equilibrium of the stage game is the profile (N, N).

Denote by $t = 1, \ldots, T$ the stage and by $\rho_i \in (0, 1]$ the discount factor of player i satisfying the condition:

$$\frac{b_i - a}{b_i - d} < \rho_i, \tag{2}$$

for any i. The above condition guarantees that if both players adopt grim trigger strategies in an infinitely repeated game, then playing (C, C) in each stage is the unique subgame-perfect Nash equilibrium.[1] Note that the above condition is always satisfied in the absence of discounting ($\rho_i = 1$).

Denote by $A = \{C, N\}$ the set of actions in the stage game, by $A_t^i \in A$ the action of player i at stage t, by $A_t = (A_t^1, A_t^2)$ the profile of players' actions at stage t, and by $u_i(A_t)$ the payoff of player $i \in I$ at stage t when the action profile A_t is played. The discounted sum of payoffs of player i is given by

$$U_i = \sum_{t=1}^{T} \rho_i^{t-1} u_i(t).$$

A history $\mathcal{H}(t)$ of the game at stage t is the sequence of realized action profiles at all stages before t, that is,

$$\mathcal{H}(t) = (A_1, \ldots, A_{t-1}), \quad \text{for } t = 2, 3, \ldots,$$

with $\mathcal{H}(1) = \varnothing$.

2.1 Limited retaliation strategy

We construct a behavior strategy that determines a player's actions in the repeated game at each stage $t + 1$ in a T-stage repeated game, based on the history $\mathcal{H}(t)$, with the assumption that players have agreed to implement some form of punishment in the case of deviation. We call this strategy the limited retaliation strategy (LRS). Whereas the well-known grim trigger strategy (GTS) prescribes to play N until the end of the game after a first defection, LRS forgives a first deviation, that is, cooperation can (possibly) resume after a noncooperative period, referred to as punishment period, during which both

[1] The inequality in (2) reflects that a one-shot deviation is not profitable, that is:

$$a + \rho_i a + \rho_i^2 a + \ldots > b_i + \rho_i d + \rho_i^2 d + \ldots \quad \Leftrightarrow \frac{\rho_i}{1 - \rho_i}(a - d) > b_i - a \Leftrightarrow \frac{b_i - a}{b_i - d} < \rho_i.$$

players play N. If a second defection occurs, then LRS and GTS coincide in prescribing to play N until the end of the game.

We introduce the following notations:

$(C,C)_{[i,k]}$: sequence of action profile (C,C) played from stage i to k inclusively;

$(N,N)_{[i,k]}$: sequence of action profile (N,N) played from stage i to k inclusively;

$(N,C)_{[t]}$: action profile at stage t where player 1 deviates while player 2 cooperates;

$(C,N)_{[t]}$: action profile at stage t where player 2 deviates while player 1 cooperates.

Remark 1. To simplify the exposition, we let the first entry in the action profile to be the action of player i and the second to be the action of player j.

The LRS strategy, denoted by $\{\phi_i(t+1,H(t))\}_{t=0}^{T-1}$, is defined as follows:

$$\phi_i(t+1,\mathcal{H}(t)) = \begin{cases} C, & \text{if } \mathcal{H}(t) = (C,C)_{[1,t]} \text{ or } \mathcal{H}(t) = \varnothing, \\ C, & \text{if } t \geq \tau + m_i(\tau) + 1 \text{ and } \tau + m_i(\tau) \leq T-1, \\ & \mathcal{H}(t) = (C,C)_{[1,\tau-1]}(N_i,C_j)_{[\tau]}(N,N)_{[\tau+1,\tau+m_i(\tau)]}(C,C)_{[\tau+m_i(\tau)+1,t]}, \\ C, & \text{if } t \geq \tau + m_j(\tau) + 1 \text{ and } \tau + m_j(\tau) \leq T-1, \\ & \mathcal{H}(t) = (C,C)_{[1,\tau-1]}(N_j,C_i)_{[\tau]}(N,N)_{[\tau+1,\tau+m_j(\tau)]}(C,C)_{[\tau+m_j(\tau)+1,t]}, \\ N, & \text{otherwise.} \end{cases} \tag{3}$$

where $m_i(\tau)$ (resp. $m_j(\tau)$) is the endogenously determined length of punishment if player i (resp. player j) deviates at stage τ. In words, the strategy ϕ_i prescribes the following behavior to a player:

1. If it is the first stage or (C,C) has been observed in all previous stages, then play C.
2. If the opponent has deviated at some stage τ (history $\mathcal{H}(\tau+1) = ((C,C)_{[1,\tau-1]}\ (C_i,N_j)_{[\tau]})$ is observed), then play N for $m_j(\tau)$ stages and returns to C at stage $\tau + m_j(\tau) + 1$.
3. If the player has deviated at some stage τ, i.e., the history $\mathcal{H}(\tau+1) = ((C,C)_{[1,\tau-1]}\ (N_i,C_j)_{[\tau]})$ is observed, then (also) play N for $m_i(\tau)$ stages and return to C at stage $\tau + m_j(\tau) + 1$.
4. Otherwise, play N.

To determine the punishment (or retaliation) period $m_i(t)$ for player i, we adopt the approach in [23]. Suppose that player i deviates from cooperation at stage t and player j does not. Then, $m_i\,(t)$ is determined by the following two rules, where SDP stands for sum of discounted payoffs:

R1. Betrayed player's SDP: Player j's SDP on $[t+m_i(t)+1,T]$ should be no less than the SDP that she could get by choosing N right after the end of the retaliation period.

R2. Deviator player's SDP: Player i's SDP on $[t,t+m_i(t)]$ should be at most equal to the SDP that she could get if (C,C) was instead played on $[t,t+m_i(t)]$.

R1 means that the betrayed player should be interested in resuming cooperation after the retaliation period. Rule R2 states that the deviator's payoff during the punishment period cannot exceed what she could get if (C,C) is implemented throughout this period.

2.2 Duration of punishment stages

Denote by $\lceil X \rceil$ the rounding up to the next integer of X, and by $\lfloor X \rfloor$ the rounding down to the previous integer of X.

Proposition 1. Suppose a first deviation in the game by player i occurs at stage t. Then, a punishment $m_i(t)$ that satisfies R1 and R2 is determined by the following inequalities:

$$\log_{\rho_i}\left(\frac{a+b_i(\rho_i-1)-d\rho_i}{a-d}\right) - 1 \leq m_i(t) \leq T - t - \log_{\rho_j}\left(\frac{a+b_j(\rho_j-1)-d\rho_j}{a-d}\right), \tag{4}$$

where

$$t \leqslant \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1,$$

and $M_k = \log_{\rho_k}\left(\frac{a+b_k(\rho_k-1)-d\rho_k}{a-d}\right)$, $k = 1, 2$.

If

$$t > \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1,$$

then, there exists no punishment that satisfies R1 and R2, and the profile (N, N) is played from $t+1$ until T.

Proof. Let $U_i[j:t]$ be player i's discounted sum of payoffs in the T-stage repeated game, where $[j:t]$ means that player j deviates at stage t, and it is the first deviation observed in the game. If the brackets are empty (denoted as $U_i[]$), then no deviations occurred in the game. The notation of U with square brackets is used only in the proof to simplify the description of the history.

If player j plays C after the end of the retaliation period until the end of the game, then her payoff in the whole game is given by

$$U_j[i:t] = \sum_{\tau=1}^{t-1} \rho_j^{\tau-1} a + c\rho_j^{t-1} + \sum_{\tau=t+1}^{t+m_i(t)} \rho_j^{\tau-1} d + \sum_{\tau=t+m_i(t)+1}^{T} \rho_j^{\tau-1} a.$$

If she plays N right after the retaliation period, i.e., at stage $t + m_i(t) + 1$, then she obtains

$$\begin{aligned} U_j[i:t, j:t+m_i(t)+1] = &\sum_{\tau=1}^{t-1} \rho_j^{\tau-1} a + c\rho_j^{t-1} + \sum_{\tau=t+1}^{t+m_i(t)} \rho_j^{\tau-1} d + \\ &+ \rho_j^{t+m_i(t)} b_j + \sum_{\tau=t+m_i(t)+2}^{T} \rho_j^{\tau-1} d. \end{aligned}$$

Now, if player i plays N at stage t (deviation), and then again N during the stages from $t+1$ to $t + m_i(t)$ (punishment), then her total payoff in the game is given by

$$U_i[i:t] = \sum_{\tau=1}^{t-1} \rho_i^{\tau-1} a + b_i\rho_i^{t-1} + \sum_{\tau=t+1}^{t+m_i(t)} \rho_i^{\tau-1} d + \sum_{\tau=t+m_i(t)+1}^{T} \rho_i^{\tau-1} a,$$

and by

$$U_i[] = \sum_{\tau=1}^{T} \rho_i^{\tau-1} a,$$

if she plays C at all stages of the game.

To determine the retaliation period $m_i(t)$, the SDP must satisfy the two rules, that is,

$$\begin{aligned} \text{R1} &: U_j[i:t] \geq U_j[i:t, j:t+m_i(t)+1], \\ \text{R2} &: U_i[] \geq U_i[i:t]. \end{aligned}$$

The first inequality can be written equivalently as

$$\begin{aligned} &\rho_j^{t+m_i(t)} a + \sum_{\tau=t+m_i(t)+2}^{T} \rho_j^{\tau-1} a \geq \rho_j^{t+m_i(t)} b_j + \sum_{\tau=t+m_i(t)+2}^{T} \rho_j^{\tau-1} d, \\ \Leftrightarrow\ &\rho_j^{t+m_i(t)} a + \frac{\rho_j^{t+m_i(t)+1} a(1-\rho_j^{T-t-m_i(t)-1})}{1-\rho_j} \geq \rho_j^{t+m_i(t)} b_j + \frac{\rho_j^{t+m_i(t)+1} d(1-\rho_j^{T-t-m_i(t)-1})}{1-\rho_j}, \end{aligned}$$

$$\Leftrightarrow \frac{(a-d)\rho_j(1-\rho_j^{T-t-m_i(t)-1})}{1-\rho_j} \geq (b_j - a),$$
$$\Leftrightarrow -1 + \frac{1-\rho_j^{T-t-m_i(t)}}{1-\rho_j} \geq \frac{b_j - a}{a-d},$$
$$\Leftrightarrow \frac{a + b_j(\rho_j - 1) - \rho_j d}{a-d} \geq \rho_j^{T-t-m_i(t)},$$
$$\Leftrightarrow T - t - \log_{\rho_j}\left(\frac{a + b_j(\rho_j - 1) - d\rho_j}{a-d}\right) \geq m_i(t).$$

Consider rule R2. The inequality $U_i[] \geq U_i[i:t]$ is equivalent to

$$b_i\rho_i^{t-1} + \frac{d\rho_i^t(1-\rho_i^{m_i(t)})}{1-\rho_i} \geq a\rho_i^{t-1} + \frac{a\rho_i^t(1-\rho_i^{m_i(t)})}{1-\rho_i},$$
$$\Leftrightarrow \frac{(a-d)\rho_i(1-\rho_i^{m_i(t)})}{1-\rho_i} \leq b_i - a,$$
$$\Leftrightarrow \rho_i - \rho_i^{m_i(t)+1} \leq \frac{(b_i - a)(1-\rho_i)}{a-d},$$
$$\Leftrightarrow \frac{a + b_i(\rho_i - 1) - \rho_i d}{a-d} \geq \rho_i^{m_i(t)+1},$$
$$\Leftrightarrow \log_{\rho_i}\left(\frac{a + b_i(\rho_i - 1) - d\rho_i}{a-d}\right) - 1 \leq m_i(t),$$

which completes the proof. □

We make four remarks on the results in the above proposition. First, if a punishment exits, it is not necessarily unique. Indeed, any $m_i(t)$ satisfying the inequality (4) is feasible. Second, the lower bound on $m_i(t)$ is independent of the stage at which the deviation has taken place, while the upper bound depends on both the stage and the terminal stage. Third, the lower bound involves (betrayal) player i's data, i.e., b_i and ρ_i while the upper bound exhibits (betrayed) player j's data, i.e., b_j and ρ_j. Finally, if the punishment duration sets at t and $t+1$ for player i are nonempty, then the maximum duration at stage t is equal to the maximum duration at stage $t+1$ minus one. To illustrate, suppose that $1 \leq m_i(t) \leq x$, then for $t+1$ we have $1 \leq m_i(t+1) \leq x-1$.

The next proposition provides the results of comparative statics on the bounds on $m_i(t)$.

Proposition 2. The lower and upper bounds on $m_i(t)$ vary as follows with the parameter values:

1. A higher deviation punishment b_i (resp. b_{3-i}), leads to larger lower-bound value (resp. smaller upper-bound value).
2. A higher punishment payoff d, leads to larger lower-bound value (resp. smaller upper-bound value).
3. A higher punishment payoff a, leads to smaller lower-bound value (resp. larger upper-bound value).
4. A higher discount factor ρ_i (resp. ρ_{3-i}), leads to smaller lower-bound value (resp. larger upper-bound value).

Proof. We have

$$M_i = \log_{\rho_i}\left(\frac{a + b_i(\rho_i - 1) - d\rho_i}{a-d}\right) = \frac{\ln\left(\frac{a+b_i(\rho_i-1)-d\rho_i}{a-d}\right)}{\ln \rho_i}.$$

The condition in (4) can be written as

$$M_i - 1 \leq m_i(t) \leq T - t - M_{3-i}.$$

Let $F_i = \frac{a+b_i(\rho_i-1)-d\rho_i}{a-d}$. Clearly, $F_i > 0$ in all admissible parameter space defined by $b_i > a > d$, $2a > b_i + c$ and $\rho_i \in (\frac{b_i-a}{b-d}, 1)$, $i = 1, 2$. The partial derivatives are given by

$$\begin{aligned}
\frac{\partial M_i}{\partial b_i} &= \frac{\rho_i - 1}{(a-d)F_i \ln \rho_i} > 0, \quad i = 1, 2, \\
\frac{\partial M_i}{\partial d} &= \frac{(a-b_i)(1-\rho_i)}{(a-d)^2 F_i \ln \rho_i} > 0, \quad i = 1, 2, \\
\frac{\partial M_i}{\partial a} &= \frac{(b_i-d)(1-\rho_i)}{(a-d)^2 F_i \ln \rho_i} < 0, \quad i = 1, 2.
\end{aligned}$$

which leads to the first three statements in the proposition.

To prove the last item of the proposition, we rewrite M_i in the following form:

$$M_i = \frac{\ln\left(\frac{a+b_i(\rho_i-1)-d\rho_i}{a-d}\right)}{\ln \rho_i} = \frac{\ln\left(1 - Q_i(1-\rho_i)\right)}{\ln(1-(1-\rho_i))},$$

where $Q_i = \frac{b_i-d}{a-d} > 1$. Since $\rho_i > \frac{b_i-a}{b-d}$, we have: $0 < Q_i(1-\rho_i) < 1$ and $Q_i(1-\rho_i) > 1-\rho_i$.

Both numerator and denominator are monotonically increasing in ρ_i because $0 < 1-Q_i(1-\rho_i) < 1$ and $0 < 1-(1-\rho_i) < 1$, and they coincide at $\rho_i = 1$, where the values of both functions reach zero (see Fig. 1). Moreover,

$$\ln(1-(1-\rho_i)) > \ln(1-Q_i(1-\rho_i)),$$

since $0 < 1 - Q_i(1-\rho_i) < \rho_i = 1-(1-\rho_i) < 1$.

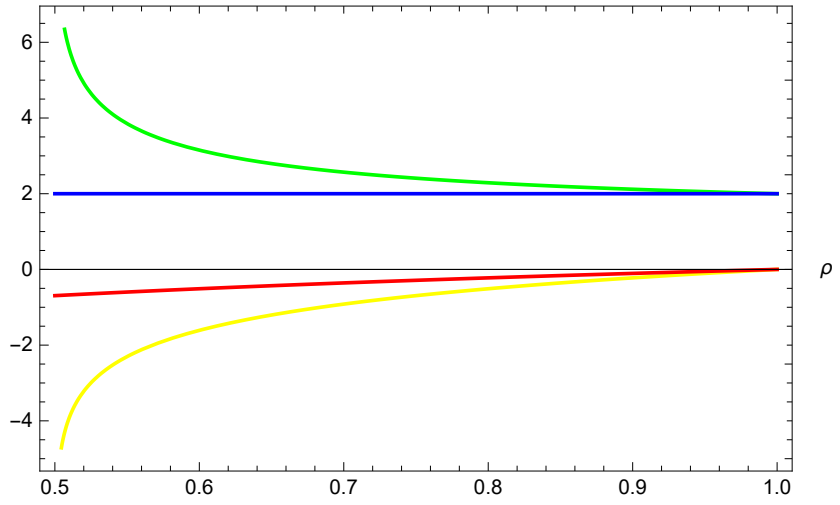

Figure 1: Graphs are constructed for $Q_i = 2$. Yellow: $\ln(1-Q_i(1-\rho_i))$, red: $\ln \rho_i$, green: M_i, blue: Q_i

Taking into account that the term $1 - Q_i(1-\rho_i)$ is increasing faster than $1-(1-\rho_i)$, then $\ln\left(1-Q_i(1-\rho_i)\right)$ is increasing slower than $\ln(1-(1-\rho_i))$, so the ratio $\ln(1-Q_i(1-\rho_i))/\ln(1-(1-\rho_i))$ is decreasing for all $\rho_i > 1 - \frac{1}{Q_i}$. Hence, we conclude that M_i monotonically decreases as ρ_i grows, with $\lim_{\rho_i \to 1} M_i = Q_i$. The limit can be calculated by L'Hôpital's rule:

$$\lim_{\rho_i \to 1} \frac{\ln\left(1-Q_i(1-\rho_i)\right)}{\ln(1-(1-\rho_i))} = \lim_{\rho_i \to 1} \frac{\left(\ln\left(1-Q_i(1-\rho_i)\right)\right)'}{\left(\ln(1-(1-\rho_i))\right)'} = \lim_{\rho_i \to 1} \frac{Q_i \rho_i}{1-Q_i(1-\rho_i)} = Q_i.$$

This finishes the proof. □

The results show that increasing the value of b_i, b_{3-i}, or d, narrows the width of the punishment interval defined in (4). On the contrary, a higher cooperative outcome a enlarges this interval. We can refine the comparative statics analysis by computing the variations in percentage and next comparing the impact of the four parameters (b_1, b_2, a, d). The elasticities are given by

$$\begin{aligned} E_{b_i}^{M_i} &= \frac{\partial M_i}{\partial b_i} \cdot \frac{b_i}{M_i} = \frac{b_i(\rho_i - 1)}{(a-d)F_i \ln F_i} > 0, \\ E_{d}^{M_i} &= \frac{\partial M_i}{\partial d} \cdot \frac{d}{M_i} = \frac{d(a-b_i)(1-\rho_i)}{(a-d)^2 F_i \ln F_i} > 0, \\ E_{a}^{M_i} &= \frac{\partial M_i}{\partial a} \cdot \frac{a}{M_i} = \frac{a(b_i-d)(1-\rho_i)}{(a-d)^2 F_i \ln F_i} < 0. \end{aligned}$$

and we have the following ratios (in absolute values):

$$\begin{aligned} \frac{|E_a^{M_i}|}{|E_d^{M_i}|} &= \frac{a(b_i-d)}{d(b_i-a)} > 1, \text{ since } a(b_i-d) > d(b_i-a) \text{ and } a > d, \\ \frac{|E_a^{M_i}|}{|E_b^{M_i}|} &= \frac{a(b_i-d)}{b_i(a-d)} > 1, \text{ since } a(b_i-d) > b_i(a-d) \text{ and } a < b_i. \end{aligned}$$

Clearly, the impact of the cooperative outcome a on the bounds in (4) is more pronounced than the impact of the deviation and punishment payoffs.

For the duration of the retaliation period to be an integer number of stages, we round the results as follows:

$$\underline{m}_i = \left\lceil M_i - 1 \right\rceil, \tag{5}$$

$$\overline{m}_i = \left\lfloor T - 1 - M_j \right\rfloor. \tag{6}$$

Observe that $\overline{m}_i = m_i(1)$, i.e., the duration of retaliation period if defection happens at stage $t = 1$. Further, from (4), we have that the maximal duration of retaliation period for any $t = 2, \ldots, T$ satisfies the following recurrence:

$$\overline{m}_i(t) = \overline{m}_i - t + 1. \tag{7}$$

Denote by t_i the last stage at which the set of punishments is nonempty for player $i = 1, 2$, i.e., this set becomes empty at $t_i + 1$. At t_i, the (rounded) minimum and maximum durations must be equal, that is,

$$\lceil M_i - 1 \rceil = \lfloor T - t_i - M_j \rfloor, \quad \text{for } i = 1, 2, \quad j = 3 - i.$$

Proposition 3. *The last stage at which the set of punishments is nonempty is defined by*

$$t_i = \left\lfloor T - M_j \right\rfloor - \left\lceil M_i \right\rceil + 1, \text{ for } i = 1, 2 \text{ and } \ j = 3 - i, \tag{8}$$

with $t_1 = t_2$.

Proof. At t_i, the minimum and maximum durations coincide, and we have

$$\begin{aligned} \underline{m}_i = \overline{m}_i(t) &= \overline{m}_i - t_i + 1 \text{ for } i \in \{1, 2\}, \\ \left\lceil M_i - 1 \right\rceil &= \left\lfloor T - 1 - M_j \right\rfloor - t_i + 1, j = 3 - i, \\ t_i &= \left\lfloor T - M_j \right\rfloor - \left\lceil M_i \right\rceil + 1. \end{aligned}$$

Compute the difference

$$\begin{aligned} t_1 - t_2 &= \left\lfloor T - M_2 \right\rfloor - \left\lceil M_1 \right\rceil + 1 - \left\lfloor T - M_1 \right\rfloor + \left\lceil M_2 \right\rceil - 1, \\ &= \left\lfloor T - M_2 \right\rfloor + \left\lceil M_2 \right\rceil - \left\lfloor T - M_1 \right\rfloor - \left\lceil M_1 \right\rceil, \\ &= T + \left\lfloor -M_2 \right\rfloor + \left\lceil M_2 \right\rceil - T - \left\lfloor -M_1 \right\rfloor - \left\lceil M_1 \right\rceil = 0, \end{aligned}$$

where we used the identity $\lfloor -x \rfloor = -\lceil x \rceil$ for any x.

Therefore, $t_1 = t_2$. □

Remark 2. If both players deviate at a same stage τ, we consider it as two deviations, and consistently with strategy (3), the players will never cooperate again. Alternatively, we could count the two simultaneous deviations as one and cooperation could resume after the end of the punishment period. In this case, the punishment length can be determined as follows: as no player benefits from the deviation, the lower bound is zero, i.e., $m(\tau) = 0$, which can be interpreted as forgiveness; the upper bound of the punishment duration is the minimum of the two lengths determined by R1. We take the minimum to ensure that R1 is satisfied for both players.

The following proposition characterizes the punishment duration in the undiscounted Prisoner's Dilemma game.

Proposition 4. Suppose that the players do not discount their stream of payoffs. The duration of the retaliation period for player i, following a deviation from cooperation at stage t and satisfying R1 and R2, is given by the solution to the following system:

$$\begin{cases} \widetilde{m_i}(t) \geq \frac{b_i - a}{a - d}, \\ \widetilde{m_i}(t) \leq T - t - 1 - \frac{b_j - a}{a - d}, \end{cases} \tag{9}$$

where

$$t \leqslant \left\lfloor T - \widetilde{M_j} \right\rfloor - \left\lceil \widetilde{M_i} \right\rceil - 1,$$

and $\widetilde{M_k} = \frac{b_k - a}{a - d}$, $k = 1, 2$.

If

$$t > \left\lfloor T - \widetilde{M_j} \right\rfloor - \left\lceil \widetilde{M_i} \right\rceil - 1,$$

then, there is no punishment satisfying R1, R2, that is, after a deviation, the players will not cooperate until the end of the game.

Proof. Player i's payoff with one deviation on step t will be:

$$U_i[i:t] = (t-1)a + b_i + \widetilde{m_i}(t)d + (T - t - \widetilde{m_i}(t))a.$$

Using Rule 2: $U_i[i:t] \leq U_i[]$, which is equivalent to the inequalities:

$$\begin{aligned} (t-1)a + b_i + \widetilde{m_i}(t)d + (T - t - \widetilde{m_i}(t))a &\leq (t-1)a + a + \widetilde{m_i}(t)a + (T - t - \widetilde{m_i}(t))a, \\ b_i + \widetilde{m_i}(t)d &\leq a + \widetilde{m_i}(t)a, \\ \frac{b_i - a}{a - d} &\leq \widetilde{m_i}(t). \end{aligned}$$

Player j's payoff when the other player deviates at stage t will be:

$$U_j[i:t] = (t-1)a + c + \widetilde{m_i}(t)d + (T - t - \widetilde{m_i}(t))a.$$

Player j's payoff when the other player deviates at stage t and after ending retaliation also decides to deviate will be:

$$U_j[i:t, j:t+\widetilde{m_i}(t)+1] = (t-1)a + c + \widetilde{m_i}(t)d + b_j + (T-t-\widetilde{m_i}(t)-1)d.$$

Using R1:

$$\begin{aligned} b_j + (T-t-\widetilde{m_i}(t)-1)d &\leq a + (T-t-\widetilde{m_i}(t)-1)a, \\ b_j - a &\leq (T-t-\widetilde{m_i}(t)-1)(a-d), \\ \frac{b_j-a}{a-d} &\leq T-t-\widetilde{m_i}(t)-1, \\ \widetilde{m_i}(t) &\leq T-t-1-\frac{b_j-a}{a-d}. \end{aligned}$$

This finishes the proof. □

Comparing the results with discounting (Proposition 1) to the ones without discounting (Proposition 4), we can notice their structural similarity. First, in both cases, the lower bound of the retaliation does not depend on either the stage at which deviation happens or the duration of the game. Second, the upper bound decreases with the stage at which a deviation happens. Despite these similarities, the no-discounting game is still of interest to assess the influence of the game duration and players' stage payoffs on sustainability of cooperation, without the effect of time preference. Further, the no-discounting case serves as a benchmark as the incentives for cooperation are the strongest possible, because future benefits are harder to ignore.

3 Subgame-perfect ε-equilibrium

For finite multi-stage games with observed actions, the one-shot deviation principle (OSDP) is a sufficient and necessary condition for subgame perfection, which directly follows from backward induction. As shown in [13], a strategy profile in a finite game of perfect information is a subgame-perfect equilibrium if and only if no player can profit by deviating from equilibrium profile in a single stage and conforming to it thereafter. This result is naturally extended to the concept of ε-equilibrium. The finiteness of the game horizon ensures that if a profitable multi-shot deviation exists, then a profitable one-shot deviation must also exist. This can be formally demonstrated by isolating the first stage at which a profitable multi-shot deviation prescribes a different action than the candidate strategy; a deviation at precisely that stage, followed by conformity, will also be profitable. Consequently, to verify that a strategy profile constitutes a subgame-perfect ε-equilibrium in a finite game, it is both necessary and sufficient to check that no player gains more than ε from any possible one-shot deviation in any subgame.

When checking if a deviation is optimal at an intermediate stage, the two continuation payoffs that must be compared can be discounted either back to the initial stage of the game or to the current stage considered as an initial stage. This difference in accounting for continuation payoffs leads to two definitions of subgame-perfect ε-equilibrium, namely, the *ex-ante* perfect ε-equilibrium (with payoffs discounted to initial stage of the game) and *contemporaneous* perfect ε-equilibrium [18]. We provide the definitions of these two subgame-perfect ε-equilibria in a T-stage repeated game and prove their existence in limited retaliation strategies.

Denote by $U_i^t\big(\sigma|H(t)\big)$ the continuation payoff to player i under strategy profile σ in the subgame starting from stage t, conditional to history $H(t)$. It is given by

$$U_i^t\big(\sigma|\mathcal{H}(t)\big) = u_i\big(A_t(\sigma|\mathcal{H}(t))\big) + \sum_{\tau=t+1}^{T} \rho_i^{\tau-t} u_i\big(A_\tau(\sigma|\mathcal{H}(t))\big) = \sum_{\tau=t}^{T} \rho_i^{\tau-t} u_i\big(A_\tau(\sigma|\mathcal{H}(t))\big), \qquad (10)$$

where $A_\tau(\sigma|\mathcal{H}(t))$ is a strategy profile at stage τ under strategy $(\sigma|H(t))$.

Definition 1. A strategy profile $\widehat{\sigma}$ in a finite T-repeated game is a contemporaneous perfect ε-equilibrium if for each player $i \in I$, every stage t, any history $\mathcal{H}(t)$, and any strategy σ_i, the following holds:

$$U_i^t(\widehat{\sigma}|\mathcal{H}(t)) \geq U_i^t(\sigma_i|\mathcal{H}(t), \widehat{\sigma}_{-i}|\mathcal{H}(t)) - \varepsilon,$$

where U_i^t is the continuation payoff of player i in the $(T-t+1)$-finitely repeated game starting at stage t with history $H(t)$ defined by equation (10), and $(\sigma_i|\mathcal{H}(t))$ is the strategy of player i in the $(T-t+1)$-finitely repeated game conditional on the history $\mathcal{H}(t)$.

If we discount the payoff values to the initial time, the strategy profile $(\sigma|\mathcal{H})$ specifies a unique history of length $t-1$ with player i's payoff computed as follows:

$$U_i(\sigma|\mathcal{H}) = \sum_{\tau=1}^{t-1} \rho_i^{\tau-1} u_i(A_\tau(\sigma|\mathcal{H})) + \rho_i^{t-1} U_i^t(\sigma|\mathcal{H}(t)), \tag{11}$$

where $U_i^t(\sigma|\mathcal{H}(t))$ is player i's payoff in the subgame starting from stage t conditional to history $\mathcal{H}(t)$ under strategy profile $(\sigma|\mathcal{H}(t))$.

Definition 2. A strategy profile $\widehat{\sigma}$ in a finite T-repeated game is an ex-ante perfect ε-equilibrium if for each player $i \in I$, any history $\mathcal{H}$, and any strategy σ_i, the following holds:

$$U_i(\widehat{\sigma}|\mathcal{H}) \geq U_i(\sigma_i|\mathcal{H}, \widehat{\sigma}_{-i}|\mathcal{H}) - \varepsilon.$$

Theorem 1. In T-repeated two-person PD game satisfying Rules R1 and R2, the profile of retaliation strategies (3) is an ex-ante perfect ε-equilibrium for any

$$\varepsilon \geq \hat{\varepsilon} = \max_{i \in I} \hat{\varepsilon}_i, \text{ where}$$
$$\hat{\varepsilon}_i = \rho_i^{T-1}(b_i - a).$$

Proof. To prove the theorem, we consider an arbitrary subgame starting at t with history $\mathcal{H}(t)$. Since the term $\sum_{\tau=1}^{t-1} \rho_i^{\tau-1} u_i(A_\tau(\sigma|\mathcal{H}))$ in formula (11) represents the stream of payoffs obtained in the periods prior to the deviation (up to stage $t-1$), its value does not vary with the strategies being compared on the right-hand and left-hand sides of the inequality in Definition 2. Consequently, this common component does not affect the result and can be eliminated in the subsequent analysis. As $\mathcal{H}(t)$ contains $t-1$ stages, the subgame will have $T-t+1$ stages. All such subgames can be partitioned into three mutually exclusive classes.

Class A (No deviation history): $\mathcal{H}(t)$ is such that both players have always followed the prescribed strategy, i.e., $\mathcal{H}(t) = (C, C)_{[1,t-1]}$.

Class B (Exactly one deviation): $\mathcal{H}(t)$ is such that in the past one deviation from mutual cooperation is observed.

Class C (Two or more deviations): $\mathcal{H}(t)$ is such that there have been at least two deviations in the past (by the same or different players).

First, consider Class C. According to R1 and R2 and LRS (3), after a second defection, the players will play noncooperatively until the end of the game. Therefore, if a player wants to deviate, she will choose C, while the other player adheres to the strategy (3). The deviator receives c instead of d, which is clearly not the best reply, i.e., a deviation is not profitable, and the Nash equilibrium will be played out at each stage of the subgame. Consequently, $\varepsilon_{C_i} = 0$ for any deviation of player i in class C.

Second, consider Class B. Since there was a deviation in this history, we have two possible situations:

1) The punishment has already ended. If any player i chooses to deviate from the limited retaliation strategy at stage $t_1 > t$, prescribing her to play C, then the other player will respond by switching to action N from next stage until the end of the game. Player i's payoff will be:

$$U_i^t[i:t_1] = \frac{a(1-\rho_i^{t_1-t})}{1-\rho_i} + \rho_i^{t_1-t} b_i + \frac{\rho_i^{t_1-t+1} d(1-\rho^{T-t_1})}{1-\rho_i}.$$

Examining $U_i^t[i:t_1]$ as a function of t_1, we obtain that it takes its maximum at $t_1 = T$. Then, the benefit from deviation is equal to

$$\varepsilon_{B1_i} = \rho_i^{t-1}(U_i^t[i:T] - U_i^t[]) = \rho_i^{t-1}\rho_i^{T-t}(b_i - a) = \rho_i^{T-1}(b_i - a).$$

2) The punishment is not over yet. Similarly to class C, it is not profitable to deviate during punishment. So if a player wants to get a positive benefit, she needs to wait for the end of the punishment and then we would move on to the case B1).

As a result, we have

$$\varepsilon_{B_i} = (b_i - a)\rho_i^{T-1}.$$

Finally, we examine Class A. It can also be divided into two points:

1) If a player has deviated only once at time $t_1 > t$, such that $t_1 \leqslant \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1$, then starting from stage $t_1 + 1$, he will receive a penalty of $m_i(t_1)$ duration and his total payoff will be

$$U_i^t[i:t_1] = \sum_{\tau=1}^{t_1-t} \rho_i^{\tau-1} a + b_i \rho_i^{t_1-t} + \sum_{\tau=t_1+2-t}^{t_1+m_i(t_1)-t+1} \rho_i^{\tau-1} d + \sum_{\tau=t_1+m_i(t_1)+2-t}^{T-t+1} \rho_i^{\tau-1} a,$$

The benefit that the player will receive in case of such a deviation, $U_i^t[i:t_1] - U_i^t[]$, will not exceed zero according to rule R2.
Thus, it is unprofitable for the player to deviate from cooperation at stage t_1, such that $t_1 \leqslant \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1$.

2) If the player deviates only once at time t_1, such that $t_1 > \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1$, then in this case the punishment of $m_i(t_1)$ duration cannot be implemented and according to strategy (3), the players switch to playing strategy N until the end of the game. The total payoff of deviator for the entire game will be equal to

$$U_i^t[i:t_1] = \frac{a(1-\rho_i^{t_1-t})}{1-\rho_i} + b_i \rho_i^{t_1-t} + \frac{d\rho_i^{t_1-t+1}(1-\rho_i^{T-t_1})}{1-\rho_i}.$$

The benefit that this player will receive in this case is:

$$\varepsilon_i = \rho_i^{t-1}\big(U_i^t[i:t_1] - U_i^t[]\big) = \rho_i^{t-1}\left((b_i - a)\rho_i^{t_1-t} - (a-d)\frac{\rho_i^{t_1-t+1}(1-\rho_i^{T-t_1})}{1-\rho_i}\right) =$$
$$= \rho_i^{t-1}\left((a-d)\frac{\rho_i^{T-t}}{1-\rho_i} + (b_i - a)\rho_i^{t_1-t} - (a-d)\frac{\rho_i^{t_1-t+1}}{1-\rho_i}\rho_i^{t-1}\right).$$

Let us find the maximum value of ε_i that player i can get by choosing a stage t_1 such that $t_1 \leq T$, $t_1 > t$ and $t_1 > \lfloor T - M_j \rfloor - \lceil M_i \rceil + 1$. The first term in the last expression does not depend on t_1 and does not affect the maximum, therefore, consider the expression

$$\rho_i^{t-1}\left((b_i - a)\rho_i^{t_1-t} - (a-d)\frac{\rho_i^{t_1-t+1}}{1-\rho_i}\right) = \frac{\rho_i^{t_1-t+t-1}}{1-\rho_i}\big((b_i - a)(1-\rho_i) - \rho_i(a-d)\big) =$$

$$= \frac{\rho_i^{t_1-1}(b_i-d)}{1-\rho_i}\left(\frac{b_i-a}{b_i-d}-\rho_i\right) \to \max_{\substack{t_1\le T,\\ t_1>\lfloor T-M_j\rfloor-\lceil M_i\rceil+1}} .$$

The multiplier $\frac{b_i-a}{b_i-d}-\rho_i<0$ since the inequality (2) holds true, then the maximum is reached at the minimum value of $\rho_i^{t_1-t}$, and since $\rho_i\in(0,1)$, then t_1 must be chosen equal to the maximum possible value, i.e., ε_i reaches the maximum value at $t_1=T$:

$$\max_{\substack{t_1\le T,\\ t_1>\lfloor T-M_j\rfloor-\lceil M_i\rceil+1}} \varepsilon_i = \rho_i^{t-1}\left((b_i-a)\rho_i^{T-t}-(a-d)\frac{\rho_i^{T-t+1}(1-\rho_i^{T-T})}{1-\rho_i}\right) =$$
$$= (b_i-a)\rho_i^{T-1}>0.$$

Thus, in this case, the condition holds with

$$\varepsilon_{A_i}=(b_i-a)\rho_i^{T-1}.$$

To conclude, by choosing

$$\hat{\varepsilon_i}=\max\{\varepsilon_{A_i},\varepsilon_{B_i},\varepsilon_{C_i}\},$$

a maximum between three values and between players, we conclude that the LRS profile is an ex-ante perfect subgame-perfect ε-equilibrium for the entire game. □

Clearly, $\hat{\varepsilon}$ is a non-decreasing function in ρ_i and b_i, $i\in I$. If both players have the same temptation payoff, i.e., $b_1=b_2$, then $\hat{\varepsilon}=\max_{i\in I}\varepsilon_i=\rho_1^{T-1}(b-a)$ if $\rho_1\ge\rho_2$. Figure 2 shows that $\hat{\varepsilon}$ is determined by the more patient player, that is, the player with higher ρ. If $\rho_1=\rho_2$, then $\varepsilon\ge\hat{\varepsilon}=\max_{i\in I}\varepsilon_i=\rho^{T-1}(b_1-a)$ when $b_1\ge b_2$. Here, $\hat{\varepsilon}$ is determined by the player who benefits more from a deviation (see Figure 3). At the point $b_1=b_2$, the graphs of $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ intersect, and any increase in asymmetry increases $\hat{\varepsilon}$. Comparing the two figures, we note that the behavior of $\hat{\varepsilon}$ after the intersection of $\hat{\varepsilon}_1$ and $\hat{\varepsilon}_2$ is less dramatic with asymmetric temptation payoffs than with asymmetric discount factors, which is due to functional form of the dependence of $\hat{\varepsilon}$ on these parameter values. In short, the maximum requirement for ε is set by the most "demanding" link: either the most patient player (who values the future more), or the one who is more tempted to deviate. As for the difference $a-d$ (gain from cooperation compared to non-cooperation), the greater it is, the harsher the retaliation for deviation, that is, the easier it is to fulfill the condition on ε.

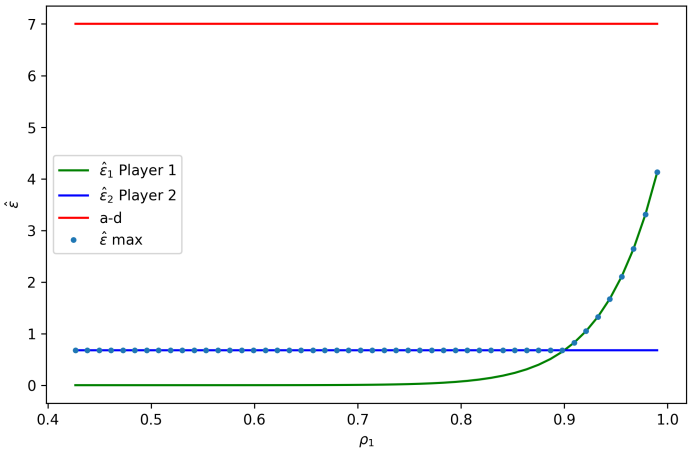

Figure 2: $\hat{\varepsilon}$ as a function of ρ_1 with fixed $\rho_2=0.9, b_1=b_2=15$. The stage payoffs are: $a=10$, $c=1$, $d=3$

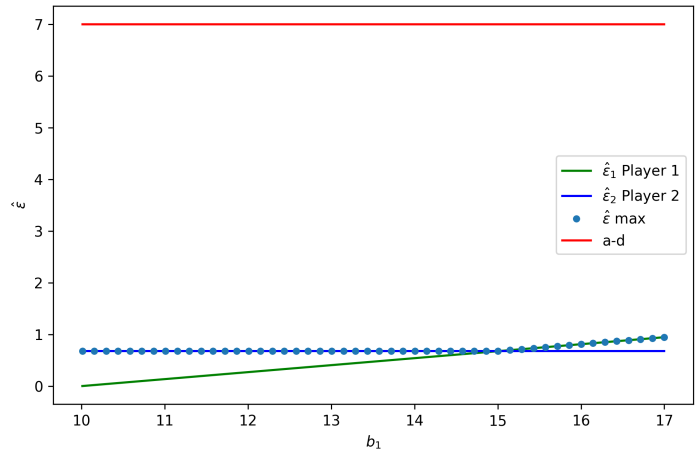

Figure 3: $\hat{\varepsilon}$ as a function of b_1 with fixed $b_2 = 15, \rho_1 = \rho_2 = 0.9$. The stage payoffs are: $a = 10$, $c = 1$, $d = 3$

Corollary 1. In finite T-repeated two-person PD game satisfying Rules R1 and R2, the profile of retaliation strategies (3) is a contemporaneous perfect ε-equilibrium for any

$$\varepsilon \geq \tilde{\varepsilon} = \max_{i \in I} \tilde{\varepsilon}_i, \text{ where}$$
$$\tilde{\varepsilon}_i = b_i - a.$$

Proof. This proof follows the same scheme as the proof of the previous theorem with one modification: here when we calculate the payoffs in the subgame $U_i^t[i : t_1]$ and $U_i^t[]$, there is no multiplier ρ_i^{t-1}. Therefore, the result is qualitatively the same but the value of $\tilde{\varepsilon}_i$ is $b_i - a$ compared to $\rho_i^{T-1}(b_i - a)$ obtained in Theorem 1. □

Clearly, ex-ante and contemporaneous perfect ε-equilibria coincide in undiscounted repeated games, with $\hat{\varepsilon}_i = \tilde{\varepsilon}_i = b_i - a$.

4 Examples

In section, we provide numerical examples to complement our analytical results on how the bounds on punishment period varies with the stage at which a deviation happens and with some parameter values, namely, b_1, b_2, ρ_1, and ρ_2.

Let $T = 20$ and the payoff matrix be given by

	C	N
C	$(10, 10)$	$(1, b_2)$
N	$(b_1, 1)$	$(3, 3)$

For each considered scenario, we determine the lower bound (LB) and upper bound (UB) for the length of the punishment phase (retaliation period) $m(t)$, and give the feasible set (FS) that contains all integer punishment durations m satisfying the inequality $LB \leq m \leq UB$. An empty set $\varnothing$ indicates that no punishment of integer length can sustain cooperation if a deviation occurs at that period.

Same temptation payoff and different discount factors. Suppose that $b_1 = b_2 = 15$, and $\rho_1 = 0.9, \rho_2 = 0.45$, that is, player 1 is more patient than player 2. With these parameter values, the pair of LR strategies are an ex-ante ε-equilibrium for

$$\varepsilon \geq \hat{\varepsilon} = \max\left\{(0.9)^{19}(15-10), (0.45)^{19}(15-10)\right\} = 0.675.$$

The results are shown in Table 1.

Table 1: Punishment duration for $b_1 = b_2 = 15, \rho_1 = 0.9, \rho_2 = 0.45$

Player 1

t	LB	UB	FS
1	0.784	15.415	$\{1, 2, \ldots, 14, 15\}$
2	0.784	14.415	$\{1, 2, \ldots, 13, 14\}$
3	0.784	13.415	$\{1, 2, \ldots, 12, 13\}$
...	...	...	...
15	0.784	1.415	$\{1\}$
16	0.784	0.415	$\varnothing$
17	0.784	-0.584	$\varnothing$
18	0.784	-1.584	$\varnothing$
19	0.784	-2.584	$\varnothing$
20	0.784	-3.584	$\varnothing$

Player 2

t	LB	UB	FS
1	2.584	17.215	$\{3, 4, \ldots, 16, 17\}$
2	2.584	16.215	$\{3, 4, \ldots, 15, 16\}$
3	2.584	15.215	$\{3, 4, \ldots, 14, 15\}$
...	...	...	...
15	2.584	3.215	$\{3\}$
16	2.584	2.215	$\varnothing$
17	2.584	1.215	$\varnothing$
18	2.584	0.215	$\varnothing$
19	2.584	-0.784	$\varnothing$
20	2.584	-0.784	$\varnothing$

We observe that the last stage at which the set of punishment period is not empty is 15. Note that the feasible sets are not identical. The constant lower bound (LB) is 0.784 for the patient player, and is much larger ($LB \approx 2.584$) for player 2.

Different temptation payoffs and same discount factor. Let $\rho_1 = \rho_2 = 0.8$, $b_1 = 16$, and $b_2 = 11$. With these parameter values, the pair of LR strategies are an ex-ante ε-equilibrium for

$$\varepsilon \geq \hat{\varepsilon} = \max\left\{(0.8)^{19}(16-10), (0.8)^{19}(11-10)\right\} = 0.08\,6$$

The results are shown in Table 2.

Table 2: Punishment duration for $\rho_1 = \rho_2 = 0.8, b_1 = 16, b_2 = 11$

Player 1

t	LB	UB	FS
1	1.080	17.837	$\{2, 3, \ldots, 16, 17\}$
2	1.080	16.837	$\{2, 3, \ldots, 15, 16\}$
3	1.080	15.837	$\{2, 3, \ldots, 14, 15\}$
...	...	...	...
15	1.080	3.837	$\{2, 3\}$
16	1.080	2.837	$\{2\}$
17	1.080	1.837	$\varnothing$
18	1.080	0.837	$\varnothing$
19	1.080	-0.162	$\varnothing$
20	1.080	-1.162	$\varnothing$

Player 2

t	LB	UB	FS
1	0.162	16.919	$\{1, 2, \ldots, 15, 16\}$
2	0.162	15.919	$\{1, 2, \ldots, 14, 15\}$
3	0.162	14.919	$\{1, 2, \ldots, 13, 14\}$
...	...	...	...
15	0.162	2.919	$\{1, 2\}$
16	0.162	1.919	$\{1\}$
17	0.162	0.919	$\varnothing$
18	0.162	-0.080	$\varnothing$
19	0.162	-1.080	$\varnothing$
20	0.162	-2.080	$\varnothing$

The main insight is that a lower temptation payoff makes cooperation easier to sustain. Indeed, player 1 needs a longer minimum punishment ($LB = 1.080$) than player 2 ($LB = 0.162$). Feasible punishment sets are nonempty for deviations up to $t = 16$. This case highlights that the magnitude of the short-term gain from deviation directly influences the required severity of punishment.

The last two examples, presented in Tables 3–4, combine asymmetries in both payoffs and discount factors.

For $b_1 = 16, b_2 = 11$ and $\rho_1 = 0.47, \rho_2 = 0.99$ (see Table 3), player 2 is extremely patient, while player 1 is not. Despite the fact that player 1 has a higher reward for temptation, player

2's high discount rate makes it possible to effectively deter deviation until $t = 13$. For player 1, the combination of low patience and high temptation severely limits cooperation, with the minimum possible punishment starts at 5.

Table 3: Punishment duration for $b_1 = 16, b_2 = 11, \rho_1 = 0.47, \rho_2 = 0.99$

Player 1

t	LB	UB	FS
1	4.5	17.85	$\{5, 6, \ldots, 16, 17\}$
2	4.5	16.85	$\{5, 6, \ldots, 15, 16\}$
...	...	...	...
13	4.5	5.85	$\{5\}$
14	4.5	4.856	$\varnothing$
15	4.5	3.856	$\varnothing$
16	4.54	2.856	$\varnothing$
17	4.5	1.856	$\varnothing$
18	4.5	0.856	$\varnothing$
19	4.5	-0.14	$\varnothing$
20	4.5	-1.14	$\varnothing$

Player 2

t	LB	UB	FS
1	0.143	13.499	$\{1, 2, \ldots, 12, 13\}$
2	0.143	12.499	$\{1, 2, \ldots, 11, 12\}$
...	...	...	...
13	0.143	1.499	$\{1\}$
14	0.143	0.499	$\varnothing$
15	0.143	-0.5	$\varnothing$
16	0.143	-1.5	$\varnothing$
17	0.143	-2.5	$\varnothing$
18	0.143	-3.5	$\varnothing$
19	0.143	-4.5	$\varnothing$
20	0.143	-6.5	$\varnothing$

When the temptation reward is higher for the patient player, that is, $b_1 = 16, b_2 = 11, \rho_1 = 0.9$, and $\rho_2 = 0.2$ (see Table 4), cooperation is sustained until stage 17.

Table 4: Punishment duration for $b_1 = 16$, $b_2 = 11$, $\rho_1 = 0.9, \rho_2 = 0.2$

Player 1

t	LB	UB	FS
1	0.949	17.473	$\{1, 2, \ldots, 16, 17\}$
2	0.949	16.473	$\{1, 2, \ldots, 15, 16\}$
...	...	...	...
13	0.949	5.473	$\{1, 2, \ldots, 4, 5\}$
14	0.949	4.856	$\{1, 2, 3, 4\}$
15	0.949	3.473	$\{1, 2, 3\}$
16	0.949	2.473	$\{1, 2\}$
17	0.949	1.473	$\{1\}$
18	0.949	0.473	$\varnothing$
19	0.949	-0.52	$\varnothing$
20	0.949	-1.52	$\varnothing$

Player 2

t	LB	UB	FS
1	0.526	17.05	$\{1, 2, \ldots, 16, 17\}$
2	0.526	16.05	$\{1, 2, \ldots, 15, 16\}$
...	...	...	...
13	0.526	5.05	$\{1, 2, 3, 4, 5\}$
14	0.526	4.05	$\{1, 2, 3, 4\}$
15	0.526	3.05	$\{1, 2, 3\}$
16	0.526	2.05	$\{1, 2\}$
17	0.526	1.05	$\{1\}$
18	0.526	0.05	$\varnothing$
19	0.526	-0.94	$\varnothing$
20	0.526	-1.94	$\varnothing$

Different temptation payoffs and no discounting. We examine asymmetry in temptations assuming $b_1 = 20$ and $b_2 = 11$. The values of LB, UB, and FS are presented in Table 5.

Table 5: Punishment duration for $b_1 = 20, b_2 = 11$, $\rho_1 = \rho_2 = 1$

Player 1

t	LB	UB	FS
1	1.429	17.857	$\{2, 3, \ldots, 16, 17\}$
2	1.429	16.857	$\{2, 3, \ldots, 15, 16\}$
...	...	...	...
13	1.429	5.857	$\{2, 3, 4, 5\}$
14	1.429	4.856	$\{2, 3, 4\}$
15	1.429	3.857	$\{2, 3\}$
16	1.429	2.857	$\{2\}$
17	1.429	1.857	$\varnothing$
18	1.429	0.857	$\varnothing$
19	1.429	-0.14	$\varnothing$
20	1.429	-1.14	$\varnothing$

Player 2

t	LB	UB	FS
1	0.143	16.571	$\{1, 2, \ldots, 15, 16\}$
2	0.143	15.571	$\{1, 2, \ldots, 14, 15\}$
...	...	...	...
13	0.143	4.571	$\{1, 2, 3, 4\}$
14	0.143	3.571	$\{1, 2, 3\}$
15	0.143	2.571	$\{1, 2\}$
16	0.143	1.571	$\{1\}$
17	0.143	0.571	$\varnothing$
18	0.143	-0.42	$\varnothing$
19	0.143	-1.42	$\varnothing$
20	0.143	-2.42	$\varnothing$

For player 1 with higher temptation $b_1 = 20$, the LB is 1.429, and for player 2 with lower temptation $b_2 = 11$, it is 0.143. Punishment is possible up to $t = 16$ inclusively, and at $t = 16$, player 1's punishment

lasts for 2 periods, while player 2's punishment only one period. Starting from $t = 17$, the FS is empty for both players.

5 Conclusions

In this paper, we demonstrate that in a finitely repeated Prisoner's Dilemma with asymmetric payoffs and discount factors, a cooperative equilibrium can be sustained using a limited retaliation strategy profile. The first result of our paper is derivation of duration of retaliation period satisfying desirable properties is that such strategies admit a perfect ε-equilibrium, providing a mechanism where mild, temporary punishment follows a first deviation, with the possibility of returning to cooperation, while a second deviation is punished until the end of the game. The second main results is the characterization of the conditions under which ex-ante and contemporaneously perfect epsilon-equilibria exist. Theoretical results of the paper are simplified on no-discounting case and demonstrated on numerical examples.

The following questions are worth considering in future research:

1. Can the examined profile of strategies become a basis for construction of payment schemes [23, 24] redistributing payments between players and in time in case of players' asymmetry?
2. How asymmetry in players' discount factors can be overcome in case of cooperation and redistribution of payoffs in time? Is it possible to construct appropriate feasible payment schemes?

References

[1] Abreu D. (1988). On the theory of infinitely repeated games with discounting. Econometrica: Journal of the Econometric Society, 383-396.

[2] Ahn, T. K., Lee, M., Ruttan, L., Walker, J. (2007). Asymmetric Payoffs in Simultaneous and Sequential Prisoner's Dilemma Games. Public Choice, 132(3/4), 353–366. http://www.jstor.org/stable/27698150

[3] Aramendia, M., Wen, Q. (2014). Justifiable punishments in repeated games, Games and Economic Behavior, 88, 16-28, https://doi.org/10.1016/j.geb.2014.07.004

[4] Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.

[5] Beckenkamp M., Hennig-Schmidt H., Maier-Rigaud F. P. (2007). Cooperation in symmetric and asymmetric prisoner's dilemma games. MPI Collective Goods Preprint. https://doi.org/10.2139/ssrn.968942

[6] Bressaud, X., Quas, A. (2017). Dynamical analysis of a repeated game with incomplete information. Mathematics of Operations Research, 42(4), 1085–1105. https://doi.org/10.1287/moor.2016.0839

[7] Chen, B., Takahashi, S. (2012). A folk theorem for repeated games with unequal discounting, Games and Economic Behavior, 76(2), 571-581, https://doi.org/10.1016/j.geb.2012.07.011

[8] Dasgupta, A., Ghosh, S. (2022). Self-accessibility and repeated games with asymmetric discounting. Journal of Economic Theory, 200, 105312, https://doi.org/10.1016/j.jet.2021.105312

[9] Ding, Y., Zhang, C., Zhang, J. (2024). Asymmetric iterated prisoner's dilemma on weighted complex networks and evolutionary strategies analysis. Journal of Statistical Mechanics: Theory and Experiment, 103402.

[10] Fong, Y.-f., Surti, J. (2009). The optimal degree of cooperation in the repeated Prisoners' Dilemma with side payments. Games and Economic Behavior. 67(1), 277-291, https://doi.org/10.1016/j.geb.2008.11.004

[11] Friedman, J. W. (1971). A non-cooperative equilibrium for supergames. 38, 1–12. Wiley-Blackwell.

[12] Fudenberg, D., Levine, D.K. (1983). Subgame-perfect equilibria of finite- and infinite-horizon games. Journal of Economic Theory, 31, 251–268.

[13] Fudenberg, D., Tirole, J. (1991). Game Theory. MIT Press.

[14] Han, Z., Zhu, P., Yang, J., Yang, J. (2023). Asymmetric players in Prisons Dilemma Game. Chaos, Solitons & Fractals, Vol. 174, 113892, https://doi.org/10.1016/j.chaos.2023.113892.

[15] Kreps, D. M., Milgrom, P., Roberts, J., Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. Journal of Economic theory, 27(2), 245–252.

[16] Lehrer, E., Pauzner, A. (1999). Repeated Games with Differential Time Preferences. Econometrica, 67, 393–412. https://doi.org/10.1111/1468-0262.00024

[17] Lehrer E., Yariv L. (1999). Repeated games with incomplete information on one side: The case of different discount factors. Mathematics of Operations Research, 24 (1), 204–218. https://doi.org/10.1287/moor.24.1.204

[18] Mailath, G.J., Postlewaite, A., Samuelson, L. (2005). Contemporaneous perfect epsilon-equilibria. Games and Economic Behavior, 53 (1), 126–140. https://doi.org/10.1016/j.geb.2005.05.002

[19] Mailath, G.J., Samuelson, L. (2006). Repeated Games and Reputations: Long-run Relationships. Oxford University Press, Oxford.

[20] Maor, C., Solan, E. (2015). Cooperation under incomplete information on the discount factors. International Journal of Game Theory, 44, 321–346.

[21] Pisareva, A.M., Parilina, E. M. (2024). Approximate Equilibrium in a Finitely Repeated Prisoner's Dilemma. Doklady Mathematics. 110(S2), 383–390.

[22] Pisareva, A. M. (2023). Construction of punishment strategy in repeated games Prisoner's Dilemma. Control Processes and Stability. 10(1). 472.

[23] Parilina E. M., Pisareva A., Zaccour G. (2025). Payment schemes for finitely repeated Prisoner's Dilemma games. Theory and Decision. 99(1-2), 461–490.

[24] Parilina, E.M., Zaccour G. (2022). Payment schemes for sustaining cooperation in dynamic games. Journal of Economic Dynamics and Control, 139, art. no. 104440.

[25] Radner R. (1980). Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. Journal of Economic Theory, 22(2), 136–154.

[26] Renault, J. (2006). The value of Markov chain games with lack of information on one side. Mathematics of Operations Research, 31 (3), 490–512. https://doi.org/10.1287/moor.1060.0199