Controlling Conditional Expectations by Zero-Determinant Strategies

Masahiko Ueda[1] ![ORCID]

Abstract

Zero-determinant strategies are memory-one strategies in repeated games which unilaterally enforce linear relations between expected payoffs of players. Recently, the concept of zero-determinant strategies was extended to the class of memory-n strategies with $n \geq 1$, which enables more complicated control of payoffs by one player. However, what we can do by memory-n zero-determinant strategies is still not clear. Here, we show that memory-n zero-determinant strategies in repeated games can be used to control conditional expectations of payoffs. Equivalently, they can be used to control expected payoffs in biased ensembles, where a history of action profiles with large value of bias function is more weighted. Controlling conditional expectations of payoffs is useful for strengthening zero-determinant strategies, because players can choose conditions in such a way that only unfavorable action profiles to one player are contained in the conditions. We provide several examples of memory-n zero-determinant strategies in the repeated prisoner's dilemma game. We also explain that a deformed version of zero-determinant strategies is easily extended to the memory-n case.

1 Introduction

Repeated games have succeeded in explaining cooperative behavior in the prisoner's dilemma situation, where defection is more favorable than cooperation [1, 2]. Recently, finite-memory strategies (strategies with finite recall) in repeated games have attracted much attention in game theory, because the rationality of real agents is bounded [3]. In computer science, agents with bounded rationality were modeled by finite automata, and equilibria of such agents have been

✉ Masahiko Ueda
  m.ueda@yamaguchi-u.ac.jp

1   Graduate School of Sciences and Technology for Innovation, Yamaguchi University,
    Yamaguchi 753-8511, Japan

investigated [4–7]. In evolutionary biology, evolutionary stability of finite-memory strategies has been mainly focused on [8–10]. The class of memory-one strategies contains several representative strategies in the repeated prisoner's dilemma game, such as the Grim Trigger strategy [11], the Tit-for-Tat strategy [12, 13], and the Win-Stay Lose-Shift strategy [9]. Moreover, longer-memory strategies have recently been investigated since longer memory enables agents more complicated behavior [14–18].

In 2012, two physicists, William Press and Freeman Dyson, discovered a novel class of memory-one strategies, called zero-determinant (ZD) strategies, in the infinitely repeated prisoner's dilemma game [19]. Counterintuitively, ZD strategies unilaterally control expected payoffs of players by enforcing linear relations between expected payoffs. Since the discovery of ZD strategies, many extensions have been done, including extensions to multi-player multi-action stage games [20–24], extensions to games with imperfect monitoring [25–27], extensions to games with a discounting factor [23, 28–30], extensions to asynchronous games [31, 32], an extension to linear relations between moments of payoffs [33], and an extension to long-memory strategies [34]. In addition, evolutionary stability of ZD strategies, such as extortionate ZD strategy and generous ZD strategy, in the repeated prisoner's dilemma game has been substantially investigated [35–40]. Human experiments also compared performance of extortionate ZD strategy and that of generous ZD strategy [41, 42]. Furthermore, mathematical properties of the situation where several players take ZD strategies were investigated [21, 27].

In this paper, we provide an interpretation about the ability of memory-n ZD strategies [34]. Although memory-n ZD strategies were originally introduced as strategies which unilaterally enforce linear relations between correlation functions of payoffs, we here elucidate that the fundamental ability of memory-n ZD strategies is that they unilaterally enforce linear relations between conditional expectations of payoffs. Equivalently, we can rephrase that memory-n ZD strategies unilaterally enforce linear relations between expected payoffs in biased ensembles [43–49]. The results in Ref. [34] can be derived from this interpretation. We also provide examples of memory-n ZD strategies in the repeated prisoner's dilemma game. Since expected payoffs conditional on previous action profiles are used in linear relations, players can choose conditions in such a way that only action profiles unfavorable to one player are contained in the conditions, which may result in strengthening original memory-one ZD strategies. Furthermore, we show that extension of deformed ZD strategies [33] to the memory-n case is straightforward.

This paper is organized as follows. In Sect. 2, we introduce a model of repeated games. In Sect. 3, we review ZD strategies. In Sect. 4, we show that there exist strategies which unilaterally enforce probability zero to specific action profiles. In Sect. 5, we introduce the concept of biased memory-n ZD strategies, and show that they unilaterally enforce linear relations between expected payoffs in biased ensembles or probability zero for a set of action profiles. In this section, we also discuss that the factorable memory-n ZD strategies in Ref. [34] can be derived from biased memory-n ZD strategies. In Sect. 6, we provide examples of biased memory-n ZD strategies in the repeated prisoner's dilemma game. In Sect. 7, we introduce memory-n version of deformed ZD strategies and provide several examples. Section 8 is devoted to concluding remarks.

2 Model

We consider a repeated game with N players. The set of players is described as $\mathcal{N} := \{1, \cdots, N\}$. The action of player $a \in \mathcal{N}$ in a one-shot game is written as $\sigma_a \in A_a := \{1, \cdots, M_a\}$, where $M_a < \infty$ is the number of action of player a. We define $\mathcal{A} := \prod_{a=1}^{N} A_a$. We collectively write $\boldsymbol{\sigma} := (\sigma_1, \cdots, \sigma_N) \in \mathcal{A}$ and call $\boldsymbol{\sigma}$ an action profile. The payoff of player a when the action profile is $\boldsymbol{\sigma}$ is described as $s_a(\boldsymbol{\sigma})$. We also write a probability M-simplex by Δ_M. We consider the situation that the game is repeated infinitely. We write an action of player a in t-th round ($t \geq 1$) by $\sigma_a(t)$. The (behavior) strategy of player a is described as $\mathcal{T}_a := \{T_a^{(t)}\}_{t=1}^{\infty}$, where $T_a^{(t)} : \mathcal{A}^{t-1} \to \Delta_{M_a}$ is the conditional probability at t-th round. We write the expectation of the quantity B with respect to strategies of all players by $\mathbb{E}[B]$. We introduce a discounting factor by δ, which satisfies $0 \leq \delta \leq 1$. The payoff of player a in the repeated game is defined by

$$\mathcal{S}_a := (1 - \delta)\mathbb{E}\left[\sum_{t=1}^{\infty} \delta^{t-1} s_a(\boldsymbol{\sigma}(t))\right] \tag{1}$$

for $0 \leq \delta < 1$, and

$$\mathcal{S}_a := \mathbb{E}\left[\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} s_a(\boldsymbol{\sigma}(t))\right] \tag{2}$$

for $\delta = 1$. In this paper, we consider only the case $\delta = 1$. Below we write $\sum_{\boldsymbol{\sigma} \in \mathcal{A}}$ and $\sum_{\sigma_a \in A_a} (\forall a)$ as $\sum_{\boldsymbol{\sigma}}$ and $\sum_{\sigma_a}$, respectively.

The payoff is rewritten as

$$\mathcal{S}_a = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{\boldsymbol{\sigma}(t)} \cdots \sum_{\boldsymbol{\sigma}(1)} s_a(\boldsymbol{\sigma}(t)) \mathbb{P}_t(\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(1)), \tag{3}$$

where we have defined the joint probability distribution of action profiles $\{\boldsymbol{\sigma}(t')\}_{t'=1}^{t}$

$$\mathbb{P}_t(\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(1)) := \prod_{t'=1}^{t} \prod_{a=1}^{N} T_a^{(t')}(\sigma_a(t')|\boldsymbol{\sigma}(t'-1), \cdots, \boldsymbol{\sigma}(1)). \tag{4}$$

It should be noted that $\mathbb{P}_t$ satisfies the recursion relation

$$\mathbb{P}_{t+1}(\boldsymbol{\sigma}(t+1), \cdots, \boldsymbol{\sigma}(1)) = \left(\prod_{a=1}^{N} T_a^{(t+1)}(\sigma_a(t+1)|\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(1))\right)\mathbb{P}_t(\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(1)). \tag{5}$$

We first introduce (time-independent) memory-n strategies ($n \geq 0$).

Definition 1 A strategy of player a is a *(time-independent) memory-n strategy* ($n \geq 0$) when it is written in the form

$$T_a^{(t)}\big(\sigma_a(t)|\boldsymbol{\sigma}(t-1),\cdots,\boldsymbol{\sigma}(1)\big) = T_a\big(\sigma_a(t)|\boldsymbol{\sigma}(t-1),\cdots,\boldsymbol{\sigma}(t-n)\big)$$
$$\left(\forall\sigma_a(t),\forall\big\{\boldsymbol{\sigma}(t')\big\}_{t'=1}^{t-1}\right) \tag{6}$$

for all $t > n$ with some common conditional probability T_a.

Throughout this paper, we consider the situation that a player $\exists a \in \mathcal{N}$ uses a memory-n strategy. We remark that strategies of players $-a := \mathcal{N}\backslash\{a\}$ are arbitrary. We define $\boldsymbol{\sigma}_{-a} := \boldsymbol{\sigma}\backslash\sigma_a$. For $t \geq n$, we also define the marginal probability distribution of the last n action profiles obtained from $\mathbb{P}_t$ by

$$P_t(\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(t-n+1)) := \sum_{\boldsymbol{\sigma}(t-n)}\cdots\sum_{\boldsymbol{\sigma}(1)}\mathbb{P}_t(\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(1)). \tag{7}$$

By taking summation of the both sides of Eq. (5) with respect to $\boldsymbol{\sigma}_{-a}(t+1)$, $\boldsymbol{\sigma}(t)$, $\cdots$, and $\boldsymbol{\sigma}(1)$ for $t \geq n$, the left-hand-side becomes

$$\sum_{\boldsymbol{\sigma}_{-a}(t+1)\in\prod_{a'\neq a}A_{a'}}\sum_{\boldsymbol{\sigma}(t)}\cdots\sum_{\boldsymbol{\sigma}(1)}\mathbb{P}_{t+1}(\boldsymbol{\sigma}(t+1),\cdots,\boldsymbol{\sigma}(1))$$
$$= \sum_{\boldsymbol{\sigma}_{-a}(t+1)\in\prod_{a'\neq a}A_{a'}}\sum_{\boldsymbol{\sigma}(t)}\cdots\sum_{\boldsymbol{\sigma}(t-n+2)}P_{t+1}(\boldsymbol{\sigma}(t+1),\cdots,\boldsymbol{\sigma}(t-n+2)) \tag{8}$$
$$= \sum_{\boldsymbol{\sigma}^{(-1)}}\cdots\sum_{\boldsymbol{\sigma}^{(-n)}}\delta_{\sigma_a^{(-1)},\sigma_a(t+1)}P_{t+1}\big(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big),$$

where $\delta_{\sigma,\sigma'}$ represents the Kronecker delta, which takes 1 for $\sigma = \sigma'$ and 0 otherwise. (The last line is obtained by renaming variables.) The right-hand-side becomes

$$\sum_{\boldsymbol{\sigma}_{-a}(t+1)\in\prod_{a'\neq a}A_{a'}}\sum_{\boldsymbol{\sigma}(t)}\cdots\sum_{\boldsymbol{\sigma}(1)}\left(\prod_{b=1}^{N}T_b^{(t+1)}\big(\sigma_b(t+1)|\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(1)\big)\right)\mathbb{P}_t(\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(1))$$
$$= \sum_{\boldsymbol{\sigma}(t)}\cdots\sum_{\boldsymbol{\sigma}(1)}T_a\big(\sigma_a(t+1)|\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(t-n+1)\big)\mathbb{P}_t(\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(1))$$
$$= \sum_{\boldsymbol{\sigma}(t)}\cdots\sum_{\boldsymbol{\sigma}(t-n+1)}T_a\big(\sigma_a(t+1)|\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(t-n+1)\big)P_t(\boldsymbol{\sigma}(t),\cdots,\boldsymbol{\sigma}(t-n+1))$$
$$= \sum_{\boldsymbol{\sigma}^{(-1)}}\cdots\sum_{\boldsymbol{\sigma}^{(-n)}}T_a\big(\sigma_a(t+1)|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big)P_t\big(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big). \tag{9}$$

By renaming $\sigma_a(t+1) \to \sigma_a$, we obtain

$$0 = \sum_{\boldsymbol{\sigma}^{(-1)}}\cdots\sum_{\boldsymbol{\sigma}^{(-n)}}T_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big)P_t\big(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big)$$
$$- \sum_{\boldsymbol{\sigma}^{(-1)}}\cdots\sum_{\boldsymbol{\sigma}^{(-n)}}\delta_{\sigma_a^{(-1)},\sigma_a}P_{t+1}\big(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big) \tag{10}$$

for $t \geq n$. Then, by calculating $\lim_{T\to\infty}\frac{1}{T}\sum_{t=n}^{T+n-1}$ of both sides, we finally obtain

$$0 = \sum_{\boldsymbol{\sigma}^{(-1)}} \cdots \sum_{\boldsymbol{\sigma}^{(-n)}} \left[T_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) - \delta_{\sigma_a^{(-1)}, \sigma_a} \right] P^*\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big), \quad (11)$$

where we have introduced the limit distribution

$$P^*\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) := \lim_{T \to \infty} \frac{1}{T} \sum_{t=n}^{T+n-1} P_t\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big). \quad (12)$$

Therefore, we obtain the generalized version of Akin's lemma [34, 50]:

Lemma 1 *For the quantity*

$$\hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) := T_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) - \delta_{\sigma_a, \sigma_a^{(-1)}}, \quad (13)$$

the relation

$$0 = \sum_{\boldsymbol{\sigma}^{(-1)}} \cdots \sum_{\boldsymbol{\sigma}^{(-n)}} \hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) P^*\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) \quad (14)$$

holds for arbitrary σ_a.

In other words, player a unilaterally enforces linear relations between values of the limit probability distribution P^* regardless of the strategies of other players. Memory-one examples of such linear relations in the repeated prisoner's game is provided in the Appendix. The quantity (13) is called a Press-Dyson tensor (or a strategy tensor) [34].

It should be noted that a Press-Dyson tensor $\hat{T}_a$ is solely controlled by player a. Due to properties of a probability distribution T_a, a Press-Dyson tensor satisfies several relations. First, it satisfies

$$\sum_{\sigma_a} \hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) = 0 \quad (15)$$

for arbitrary $\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big)$ due to the normalization condition of T_a. This implies that the number of linear relations (14) enforced by player a is at most $(M_a - 1)$. Second, it satisfies

$$\hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) \begin{cases} \leq 0, & \big(\sigma_a = \sigma_a^{(-1)}\big) \\ \geq 0, & \big(\sigma_a \neq \sigma_a^{(-1)}\big) \end{cases} \quad (16)$$

for all $\sigma_a, \boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}$. Third, it satisfies

$$\left| \hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) \right| \leq 1 \quad (17)$$

for all $\sigma_a, \boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}$. The last two comes from the fact that T_a takes value in $[0, 1]$.

Below we write the expectation for the limit distribution $P^*\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big)$ by $\langle \cdots \rangle^*$, and note $s_0(\boldsymbol{\sigma}) := 1 \ (\forall \boldsymbol{\sigma})$ for simplicity. We remark that the payoff of player $\forall a' \in \mathcal{N}$ is described as

$$
\begin{aligned}
\mathcal{S}_{a'} &= \lim_{T \to \infty} \frac{1}{T} \Bigg(\sum_{t=1}^{n-1} \sum_{\boldsymbol{\sigma}(t)} \cdots \sum_{\boldsymbol{\sigma}(1)} s_{a'}(\boldsymbol{\sigma}(t)) \mathbb{P}_t(\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(1)) \\
&\quad + \sum_{t=n}^{T} \sum_{\boldsymbol{\sigma}(t)} \cdots \sum_{\boldsymbol{\sigma}(t-n+1)} s_{a'}(\boldsymbol{\sigma}(t)) P_t(\boldsymbol{\sigma}(t), \cdots, \boldsymbol{\sigma}(t-n+1)) \Bigg) \\
&= \sum_{\boldsymbol{\sigma}^{(-1)}} \cdots \sum_{\boldsymbol{\sigma}^{(-n)}} s_{a'}\big(\boldsymbol{\sigma}^{(-1)}\big) P^*\big(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\big) \\
&= \big\langle s_{a'}\big(\boldsymbol{\sigma}^{(-1)}\big) \big\rangle^*.
\end{aligned}
\tag{18}
$$

That is, the payoffs in the repeated game are calculated as expected payoffs in the limit distribution. In the proof of Lemma 1, we have assumed that P^* exists. When P^* does not exist, the payoffs in the repeated games cannot be defined. Therefore, we consider only the case that P^* exists.

3 Previous Studies

Press and Dyson introduced the concept of zero-determinant strategies in repeated games [19]:

Definition 2 A memory-one strategy of player a is a *zero-determinant (ZD) strategy* when its Press-Dyson vectors $\hat{T}_a$ can be written in the form

$$
\sum_{\sigma_a} c_{\sigma_a} \hat{T}_a\big(\sigma_a | \boldsymbol{\sigma}^{(-1)}\big) = \sum_{b=0}^{N} \alpha_b s_b\big(\boldsymbol{\sigma}^{(-1)}\big) \quad (\forall \boldsymbol{\sigma}^{(-1)})
\tag{19}
$$

with some nontrivial coefficients $\{c_{\sigma_a}\}$ and $\{\alpha_b\}$ (that is, not $c_1 = \cdots = c_{M_a} = $ const. and not $\alpha_0 = \alpha_1 = \cdots = \alpha_N = 0$).

(Press-Dyson tensors with $n = 1$ are particularly called Press-Dyson vectors.) Because Press-Dyson vectors satisfy Akin's lemma (Lemma 1), the following proposition holds:

Proposition 1 ([19, 23]) *A ZD strategy* (19) *unilaterally enforces a linear relation between expected payoffs*:

$$
0 = \sum_{b=0}^{N} \alpha_b \big\langle s_b\big(\boldsymbol{\sigma}^{(-1)}\big) \big\rangle^*.
\tag{20}
$$

That is, the expected payoffs can be unilaterally controlled by one ZD player. Recently, a deformed version of ZD strategies was also introduced [33]:

Definition 3 A memory-one strategy of player a is a *deformed ZD strategy* when its Press-Dyson vectors $\hat{T}_a$ can be written in the form

$$\sum_{\sigma_a} c_{\sigma_a} \hat{T}_a \left(\sigma_a | \sigma^{(-1)} \right) = \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \alpha_{k_1,\cdots,k_N} \prod_{b=1}^{N} s_b \left(\sigma^{(-1)} \right)^{k_b} \quad \left(\forall \sigma^{(-1)} \right) \qquad (21)$$

with some nontrivial coefficients $\left\{ c_{\sigma_a} \right\}$ and $\left\{ \alpha_{k_1,\cdots,k_N} \right\}$.

Due to the same reason as Proposition 1, the following proposition holds:

Proposition 2 ([33]) *A deformed ZD strategy* (21) *unilaterally enforces a linear relation between moments of payoffs*:

$$0 = \sum_{k_1=0}^{\infty} \cdots \sum_{k_N=0}^{\infty} \alpha_{k_1,\cdots,k_N} \left\langle \prod_{b=1}^{N} s_b \left(\sigma^{(-1)} \right)^{k_b} \right\rangle^{*}. \qquad (22)$$

That is, the moments of payoffs can also be unilaterally controlled by one ZD player.

Furthermore, Ueda extended the concept of ZD strategies to memory-n strategies [34]:

Definition 4 A memory-n strategy of player a is a *memory-n ZD strategy* when its Press-Dyson tensors $\hat{T}_a$ can be written in the form

$$\sum_{\sigma_a} c_{\sigma_a} \hat{T}_a \left(\sigma_a | \sigma^{(-1)}, \cdots, \sigma^{(-n)} \right) = \sum_{b^{(-1)}=0}^{N} \cdots \sum_{b^{(-n)}=0}^{N} \alpha_{b^{(-1)},\cdots,b^{(-n)}} \prod_{m=1}^{n} s_{b^{(-m)}} \left(\sigma^{(-m)} \right)$$

$$\left(\forall \left\{ \sigma^{(-m)} \right\}_{m=1}^{n} \right)$$

$$(23)$$

with some nontrivial coefficients $\left\{ c_{\sigma_a} \right\}$ and $\left\{ \alpha_{b^{(-1)},\cdots,b^{(-n)}} \right\}$.

Because of Lemma 1, the following proposition also holds:

Proposition 3 ([34]) *A memory-n ZD strategy* (23) *unilaterally enforces a linear relation between correlation functions of payoffs*:

$$0 = \sum_{b^{(-1)}=0}^{N} \cdots \sum_{b^{(-n)}=0}^{N} \alpha_{b^{(-1)},\cdots,b^{(-n)}} \left\langle \prod_{m=1}^{n} s_{b^{(-m)}} \left(\sigma^{(-m)} \right) \right\rangle^{*}. \qquad (24)$$

The purpose of this paper is reinterpreting memory-n ZD strategies in terms of more elementary strategies.

4 Probability-Controlling Strategies

We first prove that there exist memory-n strategies which avoid an arbitrary action profile σ in the limit distribution. We define $\delta_{\sigma,\hat{\sigma}} := \prod_{a=1}^{N} \delta_{\sigma_a, \hat{\sigma}_a}$.

Proposition 4 *Given some action profiles* $\left\{\hat{\boldsymbol{\sigma}}^{(-m)}\right\}_{m=1}^{n}$, *memory-n strategies of player a of the form*

$$\hat{T}_a\left(\sigma_a|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right) = \hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)\prod_{m=1}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}, \tag{25}$$
$$\left(\forall\sigma_a, \forall\left\{\boldsymbol{\sigma}^{(-m)}\right\}_{m=1}^{n}\right)$$

where $\hat{T}_a^{(1)}$ *is a Press-Dyson vector of a memory-one strategy satisfying* $\hat{T}_a^{(1)}\left(\sigma_a^*|\hat{\boldsymbol{\sigma}}^{(-1)}\right) \neq 0$ *for some* σ_a^*, *unilaterally enforce probability zero to the history* $\hat{\boldsymbol{\sigma}}^{(-1)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}$:

$$P^*\left(\hat{\boldsymbol{\sigma}}^{(-1)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}\right) = 0. \tag{26}$$

Proof We consider Eq. (25) with $\sigma_a = \sigma_a^*$:

$$\hat{T}_a\left(\sigma_a^*|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right) = \hat{T}_a^{(1)}\left(\sigma_a^*|\boldsymbol{\sigma}^{(-1)}\right)\prod_{m=1}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}. \tag{27}$$

By calculating expectations of the both sides with respect to the limit distribution $P^*\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)$ corresponding to the strategy (25), and by using Lemma 1, we obtain

$$\begin{aligned}
0 &= \sum_{\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}} P^*\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)\hat{T}_a^{(1)}\left(\sigma_a^*|\boldsymbol{\sigma}^{(-1)}\right)\prod_{m=1}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}} \\
&= P^*\left(\hat{\boldsymbol{\sigma}}^{(-1)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}\right)\hat{T}_a^{(1)}\left(\sigma_a^*|\hat{\boldsymbol{\sigma}}^{(-1)}\right).
\end{aligned} \tag{28}$$

By the assumption $\hat{T}_a^{(1)}\left(\sigma_a^*|\hat{\boldsymbol{\sigma}}^{(-1)}\right) \neq 0$, we obtain the Eq. (26). □

Strategies of the form (25) can be used for avoiding some unfavorable situation $\left(\hat{\boldsymbol{\sigma}}^{(-1)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}\right)$. We call strategies of the form (25) *probability-controlling strategies*. For example, the Grim Trigger strategy of player 1 in the repeated prisoner's dilemma game can be regarded as a memory-one probability-controlling strategy avoiding the action profile (Cooperation, Defection), as we can see in the Appendix. This fact provides another explanation about the property that the Grim Trigger strategy is unbeatable [51]. We again discuss Grim Trigger in Sect. 6.

5 Biased Memory-*n* ZD Strategies

The limit distribution $P^*\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)$ gives the joint probability of *n* action profiles $\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)$. When we consider some real function $K\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)$ and introduce the quantity

$$P_K\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right) := \frac{P^*\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)e^{K\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)}}{\left\langle e^{K\left(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\right)}\right\rangle^*}, \tag{29}$$

this quantity can also be regarded as a probability distribution of n action profiles $\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right)$. In this ensemble of histories, a history with large K is more weighted. We call such ensemble P_K as *biased ensemble* biased by the function K. Biased ensembles recently attract much attention in statistical mechanics of trajectories [43–49].

We now prove our main theorem.

Theorem 1 *Let $\hat{T}_a^{(1)}$ be Press-Dyson vectors of a memory-one ZD strategy of player a*:

$$\sum_{\sigma_a} c_{\sigma_a} \hat{T}_a^{(1)}\left(\sigma_a | \sigma^{(-1)}\right) = \sum_{b=0}^{N} \alpha_b s_b\left(\sigma^{(-1)}\right) \quad \left(\forall \sigma^{(-1)}\right) \tag{30}$$

with some coefficients $\left\{c_{\sigma_a}\right\}$ and $\left\{\alpha_b\right\}$. Let $K : \mathcal{A}^n \to \mathbb{R} \cup \{-\infty\}$ be a function satisfying $K(\cdot) < \infty$, and define

$$K_{\max} := \max_{\sigma^{(-1)}, \cdots, \sigma^{(-n)}} K\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right). \tag{31}$$

Then, a memory-n strategy

$$\hat{T}_a\left(\sigma_a | \sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) = \hat{T}_a^{(1)}\left(\sigma_a | \sigma^{(-1)}\right) e^{K\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) - K_{\max}} \tag{32}$$
$$\left(\forall \sigma_a, \forall \left\{\sigma^{(-m)}\right\}_{m=1}^{n}\right)$$

unilaterally enforces either a linear relation between expected payoffs in a biased ensemble (biased by the function K)

$$0 = \sum_{b=0}^{N} \alpha_b \frac{\left\langle s_b\left(\sigma^{(-1)}\right) e^{K\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right)} \right\rangle^*}{\left\langle e^{K\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right)} \right\rangle^*} \tag{33}$$

or the relation

$$0 = P^*\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) \quad \left(\forall \left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) \in \operatorname{supp} e^{K(\cdot)}\right), \tag{34}$$

where $\operatorname{supp} f$ represents the support of function f.

Proof First, we check that tensors (32) indeed satisfy the conditions of strategies, that is, Eqs. (15), (16), and (17). Due to the equality

$$\sum_{\sigma_a} \hat{T}_a^{(1)}\left(\sigma_a | \sigma^{(-1)}\right) = 0 \quad \left(\forall \sigma^{(-1)}\right) \tag{35}$$

for Press-Dyson vectors of memory-one ZD strategies, we obtain

$$\sum_{\sigma_a} \hat{T}_a\left(\sigma_a | \sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) = \sum_{\sigma_a} \hat{T}_a^{(1)}\left(\sigma_a | \sigma^{(-1)}\right) e^{K\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) - K_{\max}} \tag{36}$$
$$= 0$$

for arbitrary $\left\{\boldsymbol{\sigma}^{(-m)}\right\}_{m=1}^{n}$, which implies Eq. (15). In addition, because the Press-Dyson vectors $\hat{T}_a^{(1)}$ of a memory-one ZD strategy satisfies

$$\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)\begin{cases} \leq 0, & \left(\sigma_a = \sigma_a^{(-1)}\right) \\ \geq 0, & \left(\sigma_a \neq \sigma_a^{(-1)}\right) \end{cases} \tag{37}$$

for all σ_a and $\boldsymbol{\sigma}^{(-1)}$, and the sign of $\hat{T}_a\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)$ is the same as that of $\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)$, we obtain Eq. (16) for all σ_a and $\left\{\boldsymbol{\sigma}^{(-m)}\right\}_{m=1}^{n}$. Furthermore, since the Press-Dyson vectors $\hat{T}_a^{(1)}$ of a memory-one ZD strategy satisfies

$$\left|\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)\right| \leq 1 \quad \left(\forall \sigma_a, \forall \boldsymbol{\sigma}^{(-1)}\right), \tag{38}$$

and then Eq. (32) satisfies

$$\begin{aligned} \left|\hat{T}_a\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)\right| &= \left|\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)-K_{\max}}\right| \\ &= \left|\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)\right|e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)-K_{\max}} \\ &\leq \left|\hat{T}_a^{(1)}\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}\right)\right| \\ &\leq 1 \quad \left(\forall \sigma_a, \forall \left\{\boldsymbol{\sigma}^{(-m)}\right\}_{m=1}^{n}\right), \end{aligned} \tag{39}$$

we obtain Eq. (17) for all $\sigma_a, \boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}$

Next, from Eqs. (32) and (30), we obtain

$$\sum_{\sigma_a} c_{\sigma_a}\hat{T}_a\left(\sigma_a|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right) = \sum_{b=0}^{N} \alpha_b s_b\left(\boldsymbol{\sigma}^{(-1)}\right)e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)-K_{\max}}. \tag{40}$$

By calculating expectations of the both sides with respect to the corresponding limit distribution $P^*\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)$ and using Lemma 1, we obtain

$$0 = \sum_{b=0}^{N} \alpha_b \left\langle s_b\left(\boldsymbol{\sigma}^{(-1)}\right)e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)-K_{\max}} \right\rangle^*. \tag{41}$$

Furthermore, if

$$\left\langle e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)} \right\rangle^* \neq 0, \tag{42}$$

by dividing the both sides of Eq. (41) by $\left\langle e^{K\left(\boldsymbol{\sigma}^{(-2)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)-K_{\max}} \right\rangle^*$, we obtain Eq. (33). Otherwise, the equality

$$\left\langle e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)} \right\rangle^* = 0 \tag{43}$$

holds. This equality is rewritten as

$$0 = \sum_{\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}} P^*\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right) e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)}. \tag{44}$$

However, because e^K is non-negative, this equality implies Eq. (34). $\qquad\square$

Theorem 1 can be regarded as an extension of Proposition 4. It should be noted that the limit probability distribution $P^*\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)$ depends on strategies. We call strategies in this Theorem as *biased memory-n ZD strategies*. Controlling biased expectations of payoffs is useful for strengthening memory-one ZD strategies, because players can choose biased functions in such a way that unfavorable action profiles to one player are more weighted. Biased ensembles are used to amplify rare events in the same way as evolution in population genetics. When we consider situation where each group with N players is selected by fitness e^K, expected payoffs in such situation are calculated by our biased expectations. Such situation may be useful in the context of multilevel selection [52], if K is given by the total payoffs of all players in one group, for instance. Furthermore, Theorem 1 contains the following three corollaries.

Corollary 1 *Let $\hat{T}_a^{(1)}$ be Press-Dyson vectors of a memory-one ZD strategy satisfying Eq. (30). Let $\left\{\hat{\boldsymbol{\sigma}}^{(-m)}\right\}_{m=1}^n$ be some action profiles. If $\sum_{b=0}^N \alpha_b s_b\left(\hat{\boldsymbol{\sigma}}^{(-1)}\right) \neq 0$, then a memory-n strategy*

$$\hat{T}_a\left(\sigma_a | \boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right) = \hat{T}_a^{(1)}\left(\sigma_a | \boldsymbol{\sigma}^{(-1)}\right) \prod_{m=1}^n \delta_{\boldsymbol{\sigma}^{(-m)}, \hat{\boldsymbol{\sigma}}^{(-m)}} \tag{45}$$
$$\left(\forall \sigma_a, \forall \left\{\boldsymbol{\sigma}^{(-m)}\right\}_{m=1}^n\right)$$

unilaterally enforces the equation

$$P^*\left(\hat{\boldsymbol{\sigma}}^{(-1)}, \cdots, \hat{\boldsymbol{\sigma}}^{(-n)}\right) = 0. \tag{46}$$

Proof By choosing the function K such that

$$e^{K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)} = \prod_{m=1}^n \delta_{\boldsymbol{\sigma}^{(-m)}, \hat{\boldsymbol{\sigma}}^{(-m)}} \tag{47}$$

in Eq. (32), we obtain

$$0 = \sum_{b=0}^N \alpha_b \left\langle s_b\left(\boldsymbol{\sigma}^{(-1)}\right) \prod_{m=1}^n \delta_{\boldsymbol{\sigma}^{(-m)}, \hat{\boldsymbol{\sigma}}^{(-m)}} \right\rangle^*. \tag{48}$$

(We remark that K can be $-\infty$.) It should be noted that $K_{\max} = 0$. By using the assumption $\sum_{b=0}^N \alpha_b s_b\left(\hat{\boldsymbol{\sigma}}^{(-1)}\right) \neq 0$, we obtain the result (46). $\qquad\square$

Corollary 1 can also be derived directly from Proposition 4.

Corollary 2 *Let $\hat{T}_a^{(1)}$ be Press-Dyson vectors of a memory-one ZD strategy satisfying Eq. (30). Let $\left\{\hat{\boldsymbol{\sigma}}^{(-m)}\right\}_{m=2}^n$ be some action profiles. Then a memory-n strategy*

$$\hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big) = \hat{T}_a^{(1)}\big(\sigma_a|\boldsymbol{\sigma}^{(-1)}\big)\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}} \tag{49}$$
$$\big(\forall\sigma_a,\forall\big\{\boldsymbol{\sigma}^{(-m)}\big\}_{m=1}^{n}\big)$$

unilaterally enforces either a linear relation between conditional expectations of payoffs when the history of the previous $n-1$ action profiles is $\hat{\boldsymbol{\sigma}}^{(-2)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}$

$$0 = \sum_{b=0}^{N}\alpha_b\frac{\big\langle s_b\big(\boldsymbol{\sigma}^{(-1)}\big)\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}\big\rangle^*}{\big\langle\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}\big\rangle^*} \tag{50}$$

or the relation

$$0 = \left\langle\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}\right\rangle^*. \tag{51}$$

Proof By choosing the function K such that

$$e^{K(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)})} = \prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}} \tag{52}$$

in Eq. (32), we obtain the relation (50) or the relation (51). It should be noted that Eq. (51) can be rewritten as

$$\begin{aligned}0 &= \sum_{\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}}P^*\big(\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big)\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}}\\ &= \sum_{\boldsymbol{\sigma}^{(-1)}}P^*\big(\boldsymbol{\sigma}^{(-1)},\hat{\boldsymbol{\sigma}}^{(-2)}\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}\big),\end{aligned} \tag{53}$$

which is a relation on the marginal distribution. $\square$

We remark that memory-n strategies (49) approach the strategy "Repeat" [50]

$$\hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big) = 0 \quad \big(\forall\sigma_a,\forall\big\{\boldsymbol{\sigma}^{(-m)}\big\}_{m=1}^{n}\big) \tag{54}$$

(which repeats the previous action of the player) as n increases, because $\prod_{m=2}^{n}\delta_{\boldsymbol{\sigma}^{(-m)},\hat{\boldsymbol{\sigma}}^{(-m)}} = 0$ for most $\big\{\boldsymbol{\sigma}^{(-m)}\big\}_{m=1}^{n}$. This property is used to control only conditional expectations when history of the action profiles is $\big(\hat{\boldsymbol{\sigma}}^{(-2)},\cdots,\hat{\boldsymbol{\sigma}}^{(-n)}\big)$.

Corollary 3 *Memory-n strategies of the form* (32) *contain factorable memory-n ZD strategies* (23):

$$\sum_{\sigma_a}c_{\sigma_a}\hat{T}_a\big(\sigma_a|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big) = \prod_{m=1}^{n}\sum_{b_m=0}^{N}\alpha_{b_m}^{(m)}s_{b_m}\big(\boldsymbol{\sigma}^{(-m)}\big). \tag{55}$$

Proof In Theorem [1], when the quantity $K\left(\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right)$ is written in the form $\sum_{m=2}^{n} K_m\left(\boldsymbol{\sigma}^{(-m)}\right)$, and the quantity $G_m(\boldsymbol{\sigma}) := e^{K_m(\boldsymbol{\sigma})} \geq 0 \ (m \geq 2)$ is written by payoffs in the form

$$G_m(\boldsymbol{\sigma}) = \sum_{b_m=0}^{N} \alpha_{b_m}^{(m)} s_{b_m}(\boldsymbol{\sigma}) \tag{56}$$

with some coefficients $\left\{ \alpha_{b_m}^{(m)} \right\}$, then the strategies [32] are reduced to memory-n ZD strategies [55]. □

6 Examples: the Prisoner's Dilemma Game

6.1 Setup

We consider the repeated prisoner's dilemma game as an example. The prisoner's dilemma game is the simplest two-player two-action game, where each player chooses cooperation (written as 1) or defection (written as 2) in each round [19]. Payoffs of two players are

$$\left(s_1(1, 1), s_1(1, 2), s_1(2, 1), s_1(2, 2) \right) = (R, S, T, P) \tag{57}$$

$$\left(s_2(1, 1), s_2(1, 2), s_2(2, 1), s_2(2, 2) \right) = (R, T, S, P) \tag{58}$$

with $T > R > P > S$. Although the Nash equilibrium of the one-shot game is $(2, 2)$, it has been known that cooperative Nash equilibria exist when this game is repeated infinitely many times and the discounting factor is large enough (folk theorem).

6.2 Biased Memory-*n* ZD Strategies

In the repeated prisoner's dilemma game, the following strategies are contained in biased memory-*n* ZD strategies of player 1:

$$\hat{T}_1\left(1|\boldsymbol{\sigma}^{(-1)}, \cdots, \boldsymbol{\sigma}^{(-n)}\right) = \prod_{m=1}^{n} G_m\left(\boldsymbol{\sigma}^{(-m)}\right) \tag{59}$$

where G_1 is a Press-Dyson vector of memory-one ZD strategies and $G_m \ (m \geq 2)$ is non-negative quantity. For example, when we assume that $2R > T + S$ and $2P < T + S$, they are

$$\begin{aligned} G_1(\cdot) \in \Big\{ &-\frac{1}{T-P}\left[s_2(\cdot) - P\right], -\frac{1}{R-S}\left[s_2(\cdot) - R\right], \frac{1}{T-S}\left[s_1(\cdot) - s_2(\cdot)\right], \\ &-\frac{1}{B}\left[s_1(\cdot) + s_2(\cdot) - (T + S)\right], \cdots \Big\} \end{aligned} \tag{60}$$

with $B := \max\{2R - (T + S), (T + S) - 2P\}$, and

$$
\begin{aligned}
G_m(\cdot) \in \Big\{ & \frac{1}{2(R - P)}\big[s_1(\cdot) + s_2(\cdot) - 2P\big], \frac{1}{2(T - S)}\big[s_1(\cdot) - s_2(\cdot) + (T - S)\big], \\
& -\frac{1}{2(T - S)}\big[s_1(\cdot) - s_2(\cdot) - (T - S)\big], \frac{1}{T - S}\big[s_1(\cdot) - S\big], \\
& -\frac{1}{T - S}\big[s_1(\cdot) - T\big], \frac{1}{T - S}\big[s_2(\cdot) - S\big], -\frac{1}{T - S}\big[s_2(\cdot) - T\big], \cdots \Big\}
\end{aligned}
\tag{61}
$$

for each $m \geq 2$. We remark that the set of G_m ($m \geq 1$) also contains quantities which cannot be represented by payoffs. A linear relation enforced by the ZD strategy is

$$
0 = \left\langle \prod_{m=1}^{n} G_m\big(\sigma^{(-m)}\big) \right\rangle^*,
\tag{62}
$$

that is, a linear relation between correlation functions of payoffs. These strategies contain all memory-two ZD strategies reported in Ref. [34].

6.3 Extension of Tit-for-Tat

For example, the memory-n strategy

$$
\begin{aligned}
\hat{T}_1\big(1|\sigma^{(-1)}, \cdots, \sigma^{(-n)}\big) &= \frac{1}{2^{n-1}(T - S)^n}\big[s_1\big(\sigma^{(-1)}\big) - s_2\big(\sigma^{(-1)}\big)\big] \prod_{m=2}^{n} \big[s_2\big(\sigma^{(-m)}\big) - s_1\big(\sigma^{(-m)}\big) + (T - S)\big] \\
&= \big[-\delta_{\sigma^{(-1)},(1,2)} + \delta_{\sigma^{(-1)},(2,1)}\big] \prod_{m=2}^{n} \Big[\tfrac{1}{2}\delta_{\sigma^{(-m)},(1,1)} + \delta_{\sigma^{(-m)},(1,2)} + \tfrac{1}{2}\delta_{\sigma^{(-m)},(2,2)}\Big]
\end{aligned}
\tag{63}
$$

is a memory-n ZD strategy, which unilaterally enforces

$$
0 = \left\langle \big\{s_1\big(\sigma^{(-1)}\big) - s_2\big(\sigma^{(-1)}\big)\big\} \prod_{m=2}^{n} \big\{s_2\big(\sigma^{(-m)}\big) - s_1\big(\sigma^{(-m)}\big) + (T - S)\big\} \right\rangle^*.
\tag{64}
$$

When $n = 1$, this strategy is reduced to Tit-for-Tat (TFT) strategy, which unilaterally enforces $\langle s_1 \rangle^* = \langle s_2 \rangle^*$ [19]. Because Eq. (63) becomes zero when $\sigma^{(-m)} = (2, 1)$ for some $m \geq 2$, Eq. (64) can be regarded as a fairness condition between two players in the situation where $(2, 1)$ was not played in the previous $n - 1$ rounds. Since the action profile $(2, 1)$ is favorable to player 1, Eq. (64) can be regarded as a fairness condition when player 1 is in an unfavorable position. Therefore, this strategy may be stronger than TFT, and whether this statement is true or not should be studied in future.

Similarly, a slightly different memory-n strategy

$$\hat{T}_1\big(1|\boldsymbol{\sigma}^{(-1)},\cdots,\boldsymbol{\sigma}^{(-n)}\big) = \frac{1}{2^{n-1}(T-S)^n}\big[s_1\big(\boldsymbol{\sigma}^{(-1)}\big) - s_2\big(\boldsymbol{\sigma}^{(-1)}\big)\big]\prod_{m=2}^{n}\big[s_1\big(\boldsymbol{\sigma}^{(-m)}\big) - s_2\big(\boldsymbol{\sigma}^{(-m)}\big) + (T-S)\big]$$

$$= \big[-\delta_{\boldsymbol{\sigma}^{(-1)},(1,2)} + \delta_{\boldsymbol{\sigma}^{(-1)},(2,1)}\big]\prod_{m=2}^{n}\Big[\frac{1}{2}\delta_{\boldsymbol{\sigma}^{(-m)},(1,1)} + \delta_{\boldsymbol{\sigma}^{(-m)},(2,1)} + \frac{1}{2}\delta_{\boldsymbol{\sigma}^{(-m)},(2,2)}\Big]$$

$$(65)$$

is also a memory-n ZD strategy, which unilaterally enforces

$$0 = \left\langle \big\{s_1\big(\boldsymbol{\sigma}^{(-1)}\big) - s_2\big(\boldsymbol{\sigma}^{(-1)}\big)\big\}\prod_{m=2}^{n}\big\{s_1\big(\boldsymbol{\sigma}^{(-m)}\big) - s_2\big(\boldsymbol{\sigma}^{(-m)}\big) + (T-S)\big\}\right\rangle^*.$$

$$(66)$$

This strategy is also reduced to TFT when $n = 1$. Because expectation in Eq. (66) is conditional on the action profiles $(1, 1)$, $(2, 1)$ and $(2, 2)$, this strategy may be weaker than TFT. We can also construct a mixed version of (63) and (65) by choosing different G_m for each $m \geq 2$.

6.4 Grim Trigger as Biased TFT

Finally, we again interpret the Grim Trigger strategy in terms of a biased memory-one ZD strategy. We remark that the Grim Trigger strategy itself is not a ZD strategy. In Eq. (45), when we set $a = 1$, $n = 1$, $\hat{\boldsymbol{\sigma}}^{(-1)} = (1, 2)$, and choose TFT as the strategy $\hat{T}_1^{(1)}$ of player 1

$$\hat{T}_1^{(1)}\big(1|\boldsymbol{\sigma}^{(-1)}\big) = -\delta_{\boldsymbol{\sigma}^{(-1)},(1,2)} + \delta_{\boldsymbol{\sigma}^{(-1)},(2,1)},$$

$$(67)$$

we obtain

$$\hat{T}_1\big(1|\boldsymbol{\sigma}^{(-1)}\big) = -\delta_{\boldsymbol{\sigma}^{(-1)},(1,2)},$$

$$(68)$$

or $T_1\big(1|\boldsymbol{\sigma}^{(-1)}\big) = \delta_{\boldsymbol{\sigma}^{(-1)},(1,1)}$, which is the Grim Trigger strategy. It should be noted that the action profile $(1, 2)$ is unfavorable for player 1. Corollary 1 claims that the strategy unilaterally enforces

$$P^*(1, 2) = 0,$$

$$(69)$$

which is consistent with Appendix. It has been known that, although a pair of TFT is not a subgame perfect equilibrium, a pair of Grim Trigger is a subgame perfect equilibrium. Therefore, this result can be interpreted as that TFT gets strengthened by a bias function $\delta_{\boldsymbol{\sigma}^{-1},(1,2)}$.

7 Memory-n Deformed ZD Strategies

We can consider a memory-n version of deformed ZD strategies (Definition 3).

Definition 5 A memory-n strategy of player a is a *deformed memory-n ZD strategy* when its Press-Dyson tensors $\hat{T}_a$ can be written in the form

$$\hat{T}_a\left(\sigma_a|\sigma^{(-1)},\cdots,\sigma^{(-n)}\right) = \prod_{m=1}^{n} G_m\left(\sigma^{(-m)}\right), \tag{70}$$

where G_1 is a deformed memory-one ZD strategy and G_m $(m \geq 2)$ are non-negative quantities.

For example, in the prisoner's dilemma game, because of the equality

$$\frac{1}{T-S}\left[s_1(\sigma) - s_2(\sigma)\right] = \frac{1}{T^k - S^k}\left[s_1(\sigma)^k - s_2(\sigma)^k\right] \quad (k \geq 1) \tag{71}$$

$$= \frac{1}{e^{hT} - e^{hS}}\left[e^{hs_1(\sigma)} - e^{hs_2(\sigma)}\right] \quad (\forall h \in \mathbb{R}) \tag{72}$$

for all σ, the following strategy can be regarded as an extension of TFT strategy:

$$\hat{T}_1\left(1|\sigma^{(-1)},\cdots,\sigma^{(-n)}\right) = \frac{1}{T-S}\left[s_1\left(\sigma^{(-1)}\right) - s_2\left(\sigma^{(-1)}\right)\right]\prod_{m=2}^{n} G_m\left(\sigma^{(-m)}\right). \tag{73}$$

A linear relation enforced by the extended TFT is

$$0 = \left\langle \left[s_1\left(\sigma^{(-1)}\right)^k - s_2\left(\sigma^{(-1)}\right)^k\right]\prod_{m=2}^{n} G_m\left(\sigma^{(-m)}\right)\right\rangle^* \quad (k \geq 1) \tag{74}$$

$$= \left\langle \left[e^{hs_1(\sigma^{(-1)})} - e^{hs_2(\sigma^{(-1)})}\right]\prod_{m=2}^{n} G_m\left(\sigma^{(-m)}\right)\right\rangle^*. \tag{75}$$

This linear relation can be interpreted as a fairness condition between two players.

In particular, because the identity

$$\frac{1}{2(T-S)}\left[s_2(\sigma) - s_1(\sigma) + (T-S)\right] = \frac{1}{2\left(e^{hT} - e^{hS}\right)}\left[e^{hs_2(\sigma)} - e^{hs_1(\sigma)} + \left(e^{hT} - e^{hS}\right)\right] \tag{76}$$

holds for arbitrary h, the strategy (63) can be rewritten as

$$\hat{T}_1\left(1|\sigma^{(-1)},\cdots,\sigma^{(-n)}\right) = \frac{1}{2^{n-1}\prod_{m=1}^{n}\left(e^{h_m T} - e^{h_m S}\right)}\left[e^{h_1 s_1(\sigma^{(-1)})} - e^{h_1 s_2(\sigma^{(-1)})}\right]$$
$$\times \prod_{m=2}^{n}\left[e^{h_m s_2(\sigma^{(-m)})} - e^{h_m s_1(\sigma^{(-m)})} + \left(e^{h_m T} - e^{h_m S}\right)\right] \tag{77}$$

with arbitrary $\{h_m\}$. Therefore, we obtain the following proposition:

Proposition 5 *The strategy* (63) *simultaneously enforces linear relations*

$$0 = \left\langle \left\{ e^{h_1 s_1(\sigma^{(-1)})} - e^{h_1 s_2(\sigma^{(-1)})} \right\} \prod_{m=2}^{n} \left\{ e^{h_m s_2(\sigma^{(-m)})} - e^{h_m s_1(\sigma^{(-m)})} + \left(e^{h_m T} - e^{h_m S} \right) \right\} \right\rangle^* \tag{78}$$

with arbitrary $\{h_m\}$.

Similarly, the strategy (65) can be rewritten as

$$\hat{T}_1\left(1|\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right) = \frac{1}{2^{n-1} \prod_{m=1}^{n} \left(e^{h_m T} - e^{h_m S} \right)} \left[e^{h_1 s_1(\sigma^{(-1)})} - e^{h_1 s_2(\sigma^{(-1)})} \right]$$
$$\times \prod_{m=2}^{n} \left[e^{h_m s_1(\sigma^{(-m)})} - e^{h_m s_2(\sigma^{(-m)})} + \left(e^{h_m T} - e^{h_m S} \right) \right] \tag{79}$$

with arbitrary $\{h_m\}$, and the following proposition holds:

Proposition 6 *The strategy* (65) *simultaneously enforces linear relations*

$$0 = \left\langle \left\{ e^{h_1 s_1(\sigma^{(-1)})} - e^{h_1 s_2(\sigma^{(-1)})} \right\} \prod_{m=2}^{n} \left\{ e^{h_m s_1(\sigma^{(-m)})} - e^{h_m s_2(\sigma^{(-m)})} + \left(e^{h_m T} - e^{h_m S} \right) \right\} \right\rangle^* \tag{80}$$

with arbitrary $\{h_m\}$.

It should be noted that the strategies (63) and (65) themselves do not depend on parameters $\{h_m\}$. Therefore, the limit probability distribution $P^*\left(\sigma^{(-1)}, \cdots, \sigma^{(-n)}\right)$ also does not depend on $\{h_m\}$. By differentiating Eq. (78) or Eq. (80) with respect to $\{h_m\}$ arbitrary times, we can obtain an infinite number of payoff relations.

8 Concluding Remarks

In this paper, we introduced the concept of biased memory-n ZD strategies (32), which unilaterally enforce linear relations between expected payoffs in biased ensembles or probability zero for a set of action profiles. Biased memory-n ZD strategies can be used to construct the original memory-n ZD strategies in Ref. [34], which unilaterally enforce linear relations between correlation functions of payoffs. From another point of view, biased memory-n ZD strategies can be regarded as extension of probability-controlling strategies introduced in Sect. 4. Furthermore, biased memory-n ZD strategies can also be used to construct ZD strategies which unilaterally enforce linear relations between conditional expectations of payoffs. Because the expectation of payoffs conditional on previous action profiles is used, players can choose conditions in such a way that only unfavorable action profiles to one player are contained in the conditions. We provided several examples of biased memory-n ZD strategies in the repeated prisoner's dilemma game. Moreover, we explained that extension of deformed ZD strategies [33] to the memory-n case is straightforward.

The significance of this study is that we provided a method to interpret every time-independent finite-memory strategy in terms of linear relations about P^* enforced by it. As we can see in Theorem 1, when a memory-n strategy is described as a biased memory-n ZD strategy, it enforces some linear relation between expected payoffs in biased ensembles or probability zero for a set of action profiles. Even if a memory-n strategy is not described as a biased memory-n ZD strategy, it can still enforce linear relations about P^* (Lemma 1). Such result may be useful for interpretation of strategies. For example, as we saw in the repeated prisoner's dilemma game, Grim Trigger can be regarded as a memory-one strategy enforcing $P^*(1, 2) = 0$, which directly represents that Grim Trigger is unbeatable. Furthermore, for any two-player symmetric potential games, the Imitate-If-Better strategy [51], which imitates the opponent's previous action if and only if it was beaten in the previous round, has the similar property as Grim Trigger in the prisoner's dilemma game [53]. That is, it unilaterally enforces a linear relation between conditional expectations

$$0 = \left\langle \left\{ s_a\left(\boldsymbol{\sigma}^{(-1)}\right) - s_{-a}\left(\boldsymbol{\sigma}^{(-1)}\right) \right\} \mathbb{I}\left(s_{-a}\left(\boldsymbol{\sigma}^{(-1)}\right) > s_a\left(\boldsymbol{\sigma}^{(-1)}\right)\right) \right\rangle^* \tag{81}$$

($\mathbb{I}$ is an indicator function), which directly means that $\langle s_a \rangle^* \geq \langle s_{-a} \rangle^*$. (Grim Trigger is a special case of the Imitate-If-Better strategy for the prisoner's dilemma game.) In this way, controlling conditional expectations can be useful when we investigate properties of finite-memory strategies.

Before ending this paper, we make two remarks. The first remark is related to the existence of biased memory-n ZD strategies. It has been known that the existence of memory-one ZD strategies is highly dependent on the stage game [27]. For example, no memory-one ZD strategies exist in the repeated rock-paper-scissors game. However, because we construct biased memory-n ZD strategies by using memory-one ZD strategies, the existence condition of biased memory-n ZD strategies is the same as that of memory-one ZD strategies used for construction. Clarifying the existence condition of memory-one ZD strategies is an important subject of future work.

The second remark is on the relation between biased memory-n ZD strategies and equilibrium strategies. Although we explained that Grim Trigger can be obtained by biasing TFT, other strategies can also be obtained. For example, in Subsect. 6.4, if we use $\hat{\boldsymbol{\sigma}}^{(-1)} = (2, 1)$ instead of $\hat{\boldsymbol{\sigma}}^{(-1)} = (1, 2)$, Eq. (68) becomes

$$\hat{T}_1\left(1|\boldsymbol{\sigma}^{(-1)}\right) = -\delta_{\boldsymbol{\sigma}^{(-1)},(1,2)}, \tag{82}$$

which unilaterally enforces $P^*(2, 1) = 0$. This strategy forms neither a subgame perfect equilibrium nor a Nash equilibrium. Therefore, the concept of biased memory-n ZD strategies itself is generally not related to equilibrium strategies. Rather, an important implication is that all time-independent finite-memory strategies can be interpreted by linear relations about P^* enforced by it, as noted above. Furthermore, in the repeated prisoner's dilemma game, memory-one Nash equilibria were characterized by using the formalism coming with memory-one ZD strategies [50]. In particular, a pair of equalizer strategies, which are one example of memory-one ZD

strategies and unilaterally set the payoff of the opponent, forms a Nash equilibrium, because each player cannot improve his/her payoff as long as the opponent uses an equalizer strategy. We would like to investigate whether our formalism can be used for characterization of memory-n Nash equilibria in the repeated prisoner's dilemma game or not in future.

Appendix

Akin's Lemma for Deterministic Memory-one Strategies

In this appendix, we provide results of Akin's lemma in the repeated prisoner's dilemma game. We use the same notation as that in Sect. 6. We write a memory-one strategy of player 1 as

$$
\boldsymbol{T}_1(1) := \begin{pmatrix} T_1(1|1,1) \\ T_1(1|1,2) \\ T_1(1|2,1) \\ T_1(1|2,2) \end{pmatrix}.
\tag{83}
$$

The number of deterministic memory-one strategies is sixteen [54]. The results of Akin's lemma are summarized in Table 1. We can see that each strategy indeed enforces a linear relation between values of the limit probability distribution P^*.

Table 1 The results of Akin's lemma

Strategy $\boldsymbol{T}_1(C)$	Name	Eq. (14)
$(1,1,1,1)^{\mathsf{T}}$	All-C	$0 = P^*(2,1) + P^*(2,2)$
$(1,1,1,0)^{\mathsf{T}}$		$0 = P^*(2,1)$
$(1,1,0,1)^{\mathsf{T}}$		$0 = P^*(2,2)$
$(1,1,0,0)^{\mathsf{T}}$	Repeat	Nothing
$(1,0,1,1)^{\mathsf{T}}$		$0 = -P^*(1,2) + P^*(2,1) + P^*(2,2)$
$(1,0,1,0)^{\mathsf{T}}$	Tit-for-Tat	$0 = -P^*(1,2) + P^*(2,1)$
$(1,0,0,1)^{\mathsf{T}}$	Win-Stay Lose-Shift	$0 = -P^*(1,2) + P^*(2,2)$
$(1,0,0,0)^{\mathsf{T}}$	Grim Trigger	$0 = -P^*(1,2)$
$(0,1,1,1)^{\mathsf{T}}$	anti-Grim Trigger	$0 = -P^*(1,1) + P^*(2,1) + P^*(2,2)$
$(0,1,1,0)^{\mathsf{T}}$	anti-Win-Stay Lose-Shift	$0 = -P^*(1,1) + P^*(2,1)$
$(0,1,0,1)^{\mathsf{T}}$	anti-Tit-for-Tat	$0 = -P^*(1,1) + P^*(2,2)$
$(0,1,0,0)^{\mathsf{T}}$		$0 = -P^*(1,1)$
$(0,0,1,1)^{\mathsf{T}}$	anti-Repeat	$0 = -P^*(1,1) - P^*(1,2) + P^*(2,1) + P^*(2,2)$
$(0,0,1,0)^{\mathsf{T}}$		$0 = -P^*(1,1) - P^*(1,2) + P^*(2,1)$
$(0,0,0,1)^{\mathsf{T}}$		$0 = -P^*(1,1) - P^*(1,2) + P^*(2,2)$
$(0,0,0,0)^{\mathsf{T}}$	All-D	$0 = -P^*(1,1) - P^*(1,2)$

Data Availability We claim that this work is a theoretical result and there is no available data source.

Declarations

Conflict of Interest The author declares no competing interests.

References

1. Fudenberg D, Tirole J (1991) Game theory. MIT Press, Massachusetts
2. Osborne MJ, Rubinstein A (1994) A course in game theory. MIT Press, Massachusetts
3. Rubinstein A (1998) Modeling bounded rationality. MIT Press, Massachusetts
4. Kalai E, Stanford W (1988) Finite rationality and interpersonal complexity in repeated games. Econometrica: Journal of the Econometric Society 397–410
5. Neyman A (1985) Bounded complexity justifies cooperation in the finitely repeated prisoners' dilemma. Econ Lett 19(3):227–229
6. Neyman A, Okada D (1999) Strategic entropy and complexity in repeated games. Games Econom Behav 29(1–2):191–223
7. Rubinstein A (1986) Finite automata play the repeated prisoner's dilemma. J Econ Theory 39(1):83–96
8. Imhof LA, Fudenberg D, Nowak MA (2005) Evolutionary cycles of cooperation and defection. Proc Natl Acad Sci 102(31):10797–10800
9. Nowak M, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature 364(6432):56–58
10. Nowak MA, Sigmund K (1992) Tit for tat in heterogeneous populations. Nature 355(6357):250–253
11. Friedman JW (1971) A non-cooperative equilibrium for supergames. Rev Econ Stud 38(1):1–12
12. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211(4489):1390–1396
13. Rapoport A, Chammah AM, Orwant CJ (1965) Prisoner's dilemma: a study in conflict and cooperation, vol 165. University of Michigan Press
14. Hilbe C, Martinez-Vaquero LA, Chatterjee K, Nowak MA (2017) Memory-n strategies of direct reciprocity. Proc Natl Acad Sci 114(18):4715–4720
15. Li J, Kendall G (2013) The effect of memory size on the evolutionary stability of strategies in iterated prisoner's dilemma. IEEE Trans Evol Comput 18(6):819–826
16. Murase Y, Baek SK (2018) Seven rules to avoid the tragedy of the commons. J Theor Biol 449:94–102
17. Murase Y, Baek SK (2020) Five rules for friendly rivalry in direct reciprocity. Sci Rep 10:16904
18. Yi SD, Baek SK, Choi JK (2017) Combination with anti-tit-for-tat remedies problems of tit-for-tat. J Theor Biol 412:1–7
19. Press WH, Dyson FJ (2012) Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. Proc Natl Acad Sci 109(26):10409–10413

20. Guo JL (2014) Zero-determinant strategies in iterated multi-strategy games. arXiv preprint arXiv: 1409.1786

21. He X, Dai H, Ning P, Dutta R (2016) Zero-determinant strategies for multi-player multi-action iterated games. IEEE Signal Process Lett 23(3):311–315

22. Hilbe C, Wu B, Traulsen A, Nowak MA (2014) Cooperation and control in multiplayer social dilemmas. Proc Natl Acad Sci 111(46):16425–16430

23. McAvoy A, Hauert C (2016) Autocratic strategies for iterated games with arbitrary action spaces. Proc Natl Acad Sci 113(13):3573–3578

24. Pan L, Hao D, Rong Z, Zhou T (2015) Zero-determinant strategies in iterated public goods game. Sci Rep 5:13096

25. Hao D, Rong Z, Zhou T (2015) Extortion under uncertainty: zero-determinant strategies in noisy games. Phys Rev E 91:052803

26. Mamiya A, Ichinose G (2019) Strategies that enforce linear payoff relationships under observation errors in repeated prisoner's dilemma game. J Theor Biol 477:63–76

27. Ueda M, Tanaka T (2020) Linear algebraic structure of zero-determinant strategies in repeated games. PLoS ONE 15(4):e0230973

28. Hilbe C, Traulsen A, Sigmund K (2015) Partners or rivals? Strategies for the iterated prisoner's dilemma. Games Econom Behav 92:41–52

29. Ichinose G, Masuda N (2018) Zero-determinant strategies in finitely repeated games. J Theor Biol 438:61–77

30. Mamiya A, Ichinose G (2020) Zero-determinant strategies under observation errors in repeated games. Phys Rev E 102:032115

31. McAvoy A, Hauert C (2017) Autocratic strategies for alternating games. Theor Popul Biol 113:13–22

32. Young RD (2017) Press-dyson analysis of asynchronous, sequential prisoner's dilemma. arXiv preprint arXiv:1712.05048

33. Ueda M (2021) Tit-for-tat strategy as a deformed zero-determinant strategy in repeated games. J Phys Soc Jpn 90(2):025002

34. Ueda M (2021) Memory-two zero-determinant strategies in repeated games. R Soc Open Sci 8(5):202186

35. Adami C, Hintze A (2013) Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. Nat Commun 4(1):1–8

36. Hilbe C, Nowak MA, Sigmund K (2013) Evolution of extortion in iterated prisoner's dilemma games. Proc Natl Acad Sci 110(17):6913–6918

37. Hilbe C, Nowak MA, Traulsen A (2013) Adaptive dynamics of extortion and compliance. PLoS ONE 8(11):1–9

38. Stewart AJ, Plotkin JB (2012) Extortion and cooperation in the prisoner's dilemma. Proc Natl Acad Sci 109(26):10134–10135

39. Stewart AJ, Plotkin JB (2013) From extortion to generosity, evolution in the iterated prisoner's dilemma. Proc Natl Acad Sci 110(38):15348–15353

40. Szolnoki A, Perc M (2014) Evolution of extortion in structured populations. Phys Rev E 89(2):022804

41. Hilbe C, Röhl T, Milinski M (2014) Extortion subdues human players but is finally punished in the prisoner's dilemma. Nat Commun 5:3976

42. Wang Z, Zhou Y, Lien JW, Zheng J, Xu B (2016) Extortion can outperform generosity in the iterated prisoner's dilemma. Nat Commun 7:11125

43. Beck C, Schögl F (1993) Thermodynamics of chaotic systems: an introduction. Cambridge University Press

44. Garrahan JP, Jack RL, Lecomte V, Pitard E, van Duijvendijk K, van Wijland F (2007) Dynamical first-order phase transition in kinetically constrained models of glasses. Phys Rev Lett 98(19):195702

45. Giardina C, Kurchan J, Peliti L (2006) Direct evaluation of large-deviation functions. Phys Rev Lett 96(12):120603

46. Jack RL, Sollich P (2010) Large deviations and ensembles of trajectories in stochastic models. Prog Theor Phys Suppl 184:304–317

47. Lecomte V, Appert-Rolland C, van Wijland F (2005) Chaotic properties of systems with Markov dynamics. Phys Rev Lett 95(1):010601

48. Nyawo PT, Touchette H (2017) A minimal model of dynamical phase transition. EPL (Europhysics Letters) 116(5):50009
49. Ueda M, Sasa SI (2015) Replica symmetry breaking in trajectories of a driven brownian particle. Phys Rev Lett 115(8):080605
50. Akin E (2016) The iterated prisoner's dilemma: good strategies and their dynamics. Ergodic Theory, Advances in Dynamical Systems. pp 77–107
51. Duersch P, Oechssler J, Schipper BC (2012) Unbeatable imitation. Games Econom Behav 76(1):88–96
52. Traulsen A, Nowak MA (2006) Evolution of cooperation by multilevel selection. Proc Natl Acad Sci 103(29):10952–10955
53. Ueda M (2022) Unbeatable tit-for-tat as a zero-determinant strategy. J Phys Soc Jpn 91(5):054804
54. Usui Y, Ueda M (2021) Symmetric equilibrium of multi-agent reinforcement learning in repeated prisoner's dilemma. Appl Math Comput 409:126370