Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Payoff control in repeated Chicken: stationary commitments and first-collision termination

M'Hamed CHERIF

mhamedcherif@gmail.com

Université Libre de Bruxelles

Research Article

Keywords: repeated Chicken, zero-determinant strategies, cooperation balance, first passage, execution error, accumulated reward

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11714013/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Payoff control in repeated Chicken: stationary commitments and first-collision termination

M'Hamed CHERIF

Lievekaai 20, 9000 Gent, Belgium

Abstract

A memory-one commitment that controls stationary payoffs need not impose the same relation on rewards accumulated before an irreversible collision. We examine this distinction in repeated Chicken with payoff ordering $T > R > S > P$. Starting from cooperation balance, we specify a baseline-P zero-determinant family and its exact probability constraints. A supporting payoff inequality establishes an attainable cooperative best response for a stated subset of Chicken games. We then apply finite cooperation-flow telescoping to encounters that stop at the first realised mutual defection. The resulting cumulative-payoff identity retains an endpoint term; independent execution errors add a cooperation-imbalance correction. Excluding only the collision payoffs differs from excluding the entire affine reward increment when the baseline is nonzero. Bounded horizons require no limiting assumption, while an unbounded version follows under finite expected collision time. Deterministic absorbing-process calculations use one main payoff tuple, $(T, R, S, P) = (5,3,1,0)$, for both continuing and stopped comparisons. Against a specified opponent at execution-error probability 0.02, a slope-two policy exceeds a unit-slope policy's stationary X payoff by 0.7477, but its stopped cumulative payoff over 50 offered rounds is lower by 2.0210. Changing the opening-memory seed reverses this stopped comparison. The analysis distinguishes stationary response incentives from collision-terminated performance without asserting learning convergence or optimality of the stopped policies.

Keywords: repeated Chicken; zero-determinant strategies; cooperation balance; first passage; execution error; accumulated reward

1. Introduction

Repeated games provide a mathematical framework for studying strategic interactions in which players repeatedly choose actions over time. Much of the classical theory focuses on long-run behaviour, where the interaction is assumed to continue indefinitely and performance is evaluated through stationary or asymptotic average payoffs. Within this setting, zero-determinant (ZD) strategies have attracted considerable attention because they allow one player, through a suitable memory-one commitment, to enforce a linear relationship between the long-run expected payoffs of both players regardless of the opponent's strategy. Since their introduction, these strategies have been extended in numerous directions and have provided important insights into unilateral payoff control in repeated interactions.

The assumption of an indefinitely continuing interaction, however, is not appropriate for many systems in which particular outcomes permanently terminate the encounter. In applications involving autonomous vehicles, communication networks, biological competition, financial systems or other safety-critical processes, certain combinations of actions correspond to irreversible failure rather than to another stage of the repeated game. Under such

circumstances, the objective is no longer the control of stationary average rewards but the control of rewards accumulated before termination. The mathematical distinction between these objectives is subtle but fundamental, since cumulative encounter rewards depend not only on the stationary dynamics of play but also on the stopping mechanism and the contribution of the terminal state.

The repeated Chicken game provides a particularly natural setting for examining this distinction. In contrast to the repeated Prisoner's Dilemma, where mutual defection represents an unfavourable but continuing outcome, mutual defection in Chicken corresponds naturally to a collision. Treating this outcome as an absorbing event changes the interpretation of strategic commitment. A payoff relation that holds under the stationary distribution of the continuing game need not remain valid when play terminates immediately after the first realised collision. Consequently, the strategic properties of memory-one commitments cannot be inferred directly from their stationary behaviour.

The present work investigates this difference by combining the theory of payoff-controlling memory-one strategies with a first-collision stopping rule. The analysis begins with the continuing repeated Chicken game and characterises an admissible family of baseline zero-determinant commitments obtained through cooperation balance. Exact feasibility conditions are established, together with a stationary payoff relation and an attainable cooperative response under explicitly stated assumptions. The analysis is then extended to encounters that terminate at the first realised mutual defection. Using a finite cooperation-flow argument, an exact identity for cumulative encounter rewards is derived. Unlike the stationary relation, this identity necessarily retains an endpoint contribution associated with the terminal state, while independent execution errors introduce an additional correction reflecting cooperation imbalance.

In Chicken, mutual defection is the worst stage outcome for both players. In a continuing repeated-game model, play can resume after that outcome. If mutual defection is instead an irreversible collision, the encounter ends and subsequent punishment or recovery cannot occur. These models assign different meanings to a commitment’s performance. A stationary mutual-defection frequency is not a first-collision probability, and a stationary payoff relation is not automatically a relation between accumulated encounter rewards.

Press and Dyson introduced zero-determinant (ZD) strategies as a way to enforce linear long-run payoff relations [1]. Cooperation balance provides a direct derivation: the focal player’s stationary probability of having cooperated equals its probability of cooperating next. Akin’s Lemma 3.1 first retains an exact finite endpoint and then obtains the long-run balance condition [2]. Ueda makes the finite-time implication explicit for general finite-action ZD strategies, obtaining bounded cumulative expected controlled rewards without assuming convergence of the Cesàro outcome distribution [3, Eqs. (12)–(16)]. The finite endpoint mechanism used below is therefore established. Our stopped identity is its collision-specific application with a predictable-attempt argument, rather than a new general telescoping principle.

Discounted autocratic strategies also retain the effect of the opening action. McAvoy and Hauert give a bounded-potential condition for enforcing linear discounted payoff relations [4]. Ichinose and Masuda study geometric continuation in two-action repeated games [5], while Govaert and Cao derive a discounted cooperation-flow identity with an explicit initial-cooperation term [6]. A

geometric continuation horizon differs from the outcome-dependent first-DD rule studied here. The policies compared in this paper are constructed for the noiseless continuing game and then held fixed; they are not redesigned to enforce an uncorrected payoff line under each new continuation rule.

Multi-state formulations provide another connection. Palasciano characterises expectation-fixing strategies in discounted finite-state games with deterministic transitions conditional on joint actions [7, Theorems 2.1 and 3.1–3.2]. McAvoy et al. give a potential condition for stochastic games with stochastic state transitions and observations [8, unnumbered theorem, Eqs. (6)–(8)]. A collision can be encoded as entry to a zero-reward cemetery state. That representation still requires a choice of reward on entry, and discounted payoff equalities do not by themselves establish integrable undiscounted stopped totals.

The noise model also matters. Observation-error ZD models condition on private signals and have error-adjusted feasibility conditions [9]. Here outcomes are observed accurately, but intended actions may be executed incorrectly. We derive the correction from the realised cooperation probability rather than substituting a private-signal formula. Once both policies are fixed memory-one rules, first-passage rewards are computed by standard absorbing Markov reward methods [10,11]. The identity itself permits a history-dependent opponent and does not depend on the numerical four-state representation.

We bring the stationary and stopped analyses into a single model. The stationary part gives the admissible baseline-P family and an attainable opponent payoff bound under a stated additional stage-payoff condition. The stopped part identifies the endpoint and separates two meanings of collision-round exclusion. Matched numerical examples then show how termination changes policy rankings. These are mathematical fixed-policy comparisons with illustrative utilities, not fitted models of behaviour or demonstrations of equilibrium selection.

2. Model, information and payoff criteria

2.1. Stage game and conditional policies

Players X and Y choose C (yield) or D (stand firm) simultaneously. Write outcomes from X's perspective in the order CC, CD, DC, DD. The strict Chicken ordering is $T > R > S > P$, with column payoff vectors

$$s_X = (R, S, T, P)^T, s_Y = (R, T, S, P)^T. (1)$$

X's memory-one vector p gives its intended cooperation probabilities after the corresponding realised outcomes. Y's vector q is indexed in Y's own-action perspective. Consequently its cooperation probabilities in the common X-state order are

$$r = (q_1, q_3, q_2, q_4)^T. (2)$$

For the four-state numerical calculations, intended action draws are conditionally independent given the preceding state. Each player's intended action is independently flipped with probability $\varepsilon \in [0,1/2]$, independently across rounds and of the intended draws. Realised outcomes are observed accurately. Effective cooperation probabilities are $a = \varepsilon\mathbf{1} + (1 - 2\varepsilon)p$ and $c = \varepsilon\mathbf{1} + (1 - 2\varepsilon)r$, giving the continuing transition matrix

$$M_{i\cdot} = (a_i c_i, a_i(1-c_i), (1-a_i)c_i, (1-a_i)(1-c_i)). (3)$$

The stopped theorem needs only X's stated conditional realised-cooperation law; it does not require Y to be memory-one. The independent two-player error model supplies the additional joint collision-hazard bound in Section 4.4.

2.2. Initialisation and two continuation rules

Let $Z_0 \sim \mu$ be a fictitious memory seed. It is not a played round, earns no reward and cannot itself be a collision. The principal seed is CC, so $\mu = (1,0,0,0)$. Policies use their conditional probabilities at that seed for round 1. A common seed does not imply a common opening mixed action. In particular, even a fictitious DD seed is followed by an actual first round; it is not an already absorbed encounter.

In the **continuing game**, actual DD outcomes do not stop play. A stationary distribution is a row vector v with $vM = v$ and $v\mathbf{1} = 1$, and stationary payoffs are $\pi_X = vs_X$, $\pi_Y = vs_Y$. Positive error makes M strictly positive. At zero error several recurrent classes can occur. We then use the seed-selected expected Cesàro occupancy, obtained by weighting recurrent-class stationary distributions by their absorption probabilities from μ. Stationary identities below hold for every stationary distribution, without requiring uniqueness.

In a **collision-terminated encounter**, define

$$\tau = \inf\{t \geq 1 : Z_t = DD\}, L_H = \tau \wedge H, C_H = \Pr(\tau \leq H), (4)$$

where $\tau = \infty$ if no collision occurs. The offered horizon $H \geq 1$ is deterministic. No further round is attempted after an actual DD. Survival is $S_H = \Pr(\tau > H) = 1 - C_H$, and the number of safe rounds is $N_H = L_H - \mathbf{1}_{\{\tau \leq H\}}$.

2.3. Included and excluded rewards

The included expected encounter rewards are

$$U_i^{\text{inc}}(H) = \mathbb{E}\left[\sum_{t=1}^{L_H} s_i(Z_t)\right], i \in \{X, Y\}. (5)$$

The excluded reward omits the terminal DD payoff, if present, but leaves the attempted duration unchanged. Because each player's DD payoff is P,

$$U_i^{\text{inc}}(H) - U_i^{\text{exc}}(H) = PC_H. (6)$$

Excluding the entire affine reward increment will be distinguished from this reward-only exclusion in Section 4.2. All totals are unconditional expectations, not payoffs conditioned on survival. Dividing by the common offered horizon H preserves a ranking of totals; it does not give the expected within-encounter average $\mathbb{E}[G_i/L_H]$ or a stationary average.

For random-duration totals, the utility origin is substantive. Adding a constant d to every included stage reward changes a policy's total by $d\mathbb{E}L_H$. Policies with different durations can therefore change rank. The principal choice $P = 0$ is part of the specified cardinal utilities, not a harmless normalisation for arbitrary stopped-total objectives.

3. Continuing-game commitments

3.1. Cooperation balance and the admissible family

Let $e = (1,1,0,0)^T$ indicate X's realised cooperation. For a baseline b, slope $\chi \geq 1$ and scale $\phi > 0$, consider the intended policy

$$f_b = s_X - \chi s_Y + (\chi - 1)b\mathbf{1}, p = e + \phi f_b. (7)$$

Only choices with $p \in [0,1]^4$ are used. The principal family has $b = P$.

Proposition 1 (baseline-P admissibility and stationary enforcement). For strict Chicken and $\chi \geq 1$, a positive scale in (7) with $b = P$ is admissible if and only if

$$1 \leq \chi \leq \frac{T-P}{S-P}, 0 < \phi \leq \frac{1}{\max\{A_\chi, B_\chi, C_\chi\}}, (8)$$

where

$$A_\chi = (\chi - 1)(R - P), B_\chi = \chi(T - P) - (S - P), C_\chi = (T - P) - \chi(S - P). (9)$$

The policy is

$$p = (1 - \phi A_\chi, 1 - \phi B_\chi, \phi C_\chi, 0)^T. (10)$$

For $\varepsilon = 0$, every stationary distribution satisfies

$$\pi_X - P = \chi(\pi_Y - P). (11)$$

Proof. For $\chi \geq 1$, $A_\chi \geq 0$ and $B_\chi > 0$. The third probability is nonnegative at positive scale precisely when $C_\chi \geq 0$, giving the slope ceiling in (8). The remaining probability constraints are exactly $\phi A_\chi \leq 1$, $\phi B_\chi \leq 1$ and $\phi C_\chi \leq 1$. At stationarity and zero error, $vp = ve$, since both are X's cooperation probability one round apart. Thus $0 = v(p - e) = \phi v f_P$, proving (11). This is the cooperation-balance construction [1,2]. ▫

The unit-slope case $\chi = 1$ gives stationary payoff equality. For $\chi > 1$, (11) fixes a surplus ratio whenever the opponent's surplus is positive; if both surpluses vanish, it does not describe a positive advantage. Zero scale is excluded because $p = e$ alone does not imply (11). Valid probabilities do not imply irreducibility. At the upper slope endpoint, $p_{DC} = p_{DD} = 0$.

For this family, $\chi_{\max} = 1 + (T - S)/(S - P)$ and $d\chi_{\max}/dP = (T - S)/(S - P)^2 > 0$ when T, S are fixed. Lowering P narrows the admissible slope range. This is a probability constraint, not a behavioural threshold. It concerns a reconstructed baseline-P policy, not the fixed-policy terminal-charge experiment in Section 5.4.

3.2. An attainable cooperative stationary response

Stationary payoff control does not itself identify a best response. A separate payoff inequality gives one under an additional condition.

Proposition 2 (attainable stationary opponent bound). Suppose $T + S \leq 2R$ and X uses an admissible policy (10). At zero execution error, every stationary payoff pair obeys

$$\pi_Y - P \leq \frac{(R-P)(T-S)}{\chi(R-S)+(T-R)}. (12)$$

The always-cooperate opponent attains the bound.

Proof. The statewise inequality

$$(R-S)s_X + (T-R)s_Y \leq R(T-S)\mathbf{1} (13)$$

is an equality at CC and DC. At CD it is equivalent to $T+S \leq 2R$, and at DD it follows from $P < R$. Average (13) and substitute (11) to obtain (12).

Against always cooperation, after the first actual round only CC and DC can occur. The transition probabilities from CC to DC and from DC to CC are ϕA_χ and ϕC_χ. Their stationary weights are

$$v_{CC} = \frac{C_\chi}{A_\chi + C_\chi}, v_{DC} = \frac{A_\chi}{A_\chi + C_\chi}, v_{CD} = v_{DD} = 0. (14)$$

Here $A_\chi + C_\chi = \chi(R-S) + (T-R) > 0$. The formula also covers $\chi = 1$ and the upper slope endpoint: respectively CC and DC is the absorbing class against this opponent. Since (13) is an equality on the support of (14), the upper bound is attained. ▫

This result concerns the continuing, noiseless stationary objective. It makes no claim that an opponent learns to cooperate, that always cooperation is the unique conditional policy attaining the bound, or that it maximises stopped rewards. If expected Cesàro payoff limits exist against a history-dependent opponent, deterministic finite cooperation-flow balance also gives (11) for those limits, so the same supporting inequality bounds them. Always cooperation supplies the stated attainable stationary response.

For the principal tuple $(5,3,1,0)$, (8) reduces to $1 \leq \chi \leq 5$ and $0 < \phi \leq 1/(5\chi - 1)$, while (12) becomes $\pi_Y \leq 6/(\chi+1)$. At $\chi = 2$ it is attained by $v = (1/2,0,1/2,0)$, with $(\pi_X, \pi_Y) = (4,2)$. For comparison, the different tuple $(4,3,1,0)$ gives $\pi_Y \leq 9/(2\chi+1)$. We do not mix those tuples in the numerical comparisons below.

4. Cooperation flow stopped at the first collision

4.1. Bounded stopped identity

We now allow any baseline b for which (7) is an admissible policy. Let $\mathcal{F}_t$ contain the initial seed and realised history through round t. Y may use arbitrary behavioural dependence on that history. Assume X's realised cooperation satisfies

$$\mathbb{E}[e(Z_t) \mid \mathcal{F}_{t-1}] = a(Z_{t-1}) = \varepsilon + (1-2\varepsilon)p(Z_{t-1}) (15)$$

on each attempted round. A hypothetical continuation after stopping can be defined for notation; it does not contribute rewards. Write $D_H = \mathbb{E}L_H$ and $K_H = \sum_{t=1}^{L_H} e(Z_t)$, and define

$$F_H^{\mathrm{inc}} = U_X^{\mathrm{inc}} - \chi U_Y^{\mathrm{inc}} + (\chi-1)bD_H. (16)$$

Proposition 3 (bounded stopped cooperation flow). Under (7) and (15), for every deterministic $H \geq 1$,

$$(1-2\varepsilon)\phi F_H^{\mathrm{inc}} + \varepsilon(D_H - 2\mathbb{E}K_H) = (1-2\varepsilon)\{\mathbb{E}p(Z_{L_H}) - \mu p\}. \quad (17)$$

For $0 \le \varepsilon < 1/2$ this is equivalently

$$F_H^{\mathrm{inc}} = \frac{\mathbb{E}p(Z_{L_H}) - \mu p}{\phi} - \frac{\varepsilon(D_H - 2\mathbb{E}K_H)}{(1-2\varepsilon)\phi}. \quad (18)$$

Proof. Define the mean-zero increments $\Delta_t = e(Z_t) - a(Z_{t-1})$. The attempt event $\{L_H \ge t\}$ is $\mathcal{F}_{t-1}$-measurable. Boundedness of L_H therefore gives

$$\mathbb{E}\sum_{t=1}^{L_H} \Delta_t = 0. \quad (19)$$

Pathwise telescoping and (19) imply

$$\mathbb{E}\sum_{t=1}^{L_H} \{a(Z_t) - e(Z_t)\} = \mathbb{E}\{a(Z_{L_H}) - a(Z_0)\}. \quad (20)$$

Since $p - e = \phi f_b$,

$$a - e = (1-2\varepsilon)\phi f_b + \varepsilon(\mathbf{1} - 2e). \quad (21)$$

Substitute (21) into (20). The expected accumulated f_b is (16), and the constant part of a cancels at the endpoints. This proves (17); division proves (18) on its stated domain. ▫

At zero error, (18) has only the endpoint term and $|F_H^{\mathrm{inc}}| \le 1/\phi$. For deterministic, nonstopped play it is an arrival-indexed form of Akin's finite cooperation telescope and Ueda's potential identity [2,3]. Their familiar cooperation endpoint uses $\mathbb{E}e(Z_{H+1}) - \mathbb{E}e(Z_1)$; here $\mathbb{E}p(Z_H) - \mu p$ is equal to that expression at zero error. The stopping argument is needed because replacing a deterministic horizon by a random variable inside a previously averaged identity is not justified on its own.

Importantly, $\tau - 1$ is generally not a stopping time in the outcome filtration. Selecting a safe-round reward uses $\{\tau > t\}$, which is not generally known before round t's actions. We therefore stop at the attempted collision round and subtract its reward afterwards. Evaluating p at the terminal DD is a potential evaluation; it does not assert that another action is played after collision.

4.2. Two meanings of collision-round exclusion

With attempted duration retained, define

$$F_H^{\mathrm{exc,attempt}} = U_X^{\mathrm{exc}} - \chi U_Y^{\mathrm{exc}} + (\chi - 1)bD_H.$$

By (6),

$$F_H^{\mathrm{exc,attempt}} = F_H^{\mathrm{inc}} + (\chi - 1)PC_H. \quad (22)$$

If the *entire affine increment* is removed, use the expected safe duration instead:

$$F_H^{\mathrm{exc,safe}} = U_X^{\mathrm{exc}} - \chi U_Y^{\mathrm{exc}} + (\chi - 1)b\mathbb{E}N_H = F_H^{\mathrm{inc}} + (\chi - 1)(P - b)C_H. \quad (23)$$

Indeed $f_b(DD) = (1-\chi)(P-b)$, and (23) subtracts this whole increment. At $b = P$, the whole DD increment is zero, so (23) equals F_H^{inc}. Equation (22) need not. Neither convention generally deletes the terminal potential in (17). For $P = b = 0$ both corrections vanish, which makes the principal numerical example insufficient by itself to check this distinction.

A nonzero-payoff diagnostic makes the difference explicit. Set $(T,R,S,P) = (5,3,1,-1)$, $b = P = -1$, $\chi = 2$, $\phi = 0.05$, so $p = (0.8,0.5,0.1,0)$. From a fictitious CC seed against ALLD, at $\varepsilon = 0$ and $H = 1$, CD has probability 0.8 and DD probability 0.2. Included rewards are $(0.6,3.8)$ and excluded rewards $(0.8,4)$. The attempted and safe durations have expectations 1 and 0.8. Thus

$$F_1^{\mathrm{inc}} = -8, F_1^{\mathrm{exc,attempt}} = -8.2, F_1^{\mathrm{exc,safe}} = -8. \quad (24)$$

The retained boundary is $(0.4-0.8)/0.05 = -8$. The DD payoff has been excluded in both excluded quantities, but the baseline contributions differ.

4.3. Unbounded first-collision encounters

Corollary 3.1 (integrable stopping). Assume (7), (15), bounded stage payoffs and $\mathbb{E}\tau < \infty$. Equations (17), (22) and (23) extend to $L = \tau$ by taking $H \to \infty$. At zero error and baseline $b = P$,

$$U_X^{\mathrm{inc}}(\infty) - \chi U_Y^{\mathrm{inc}}(\infty) + (\chi-1)P\mathbb{E}\tau = -\frac{\mu p}{\phi}. \quad (25)$$

Proof. Each absolute reward sum and cooperation count through L_H is bounded by a constant times τ, while the endpoint potential is bounded. Dominated convergence applies. Finite mean implies almost sure collision, hence $Z_\tau = DD$ and $p_{DD} = 0$ for (10), giving (25). ▫

Finite expected time is a sufficient condition, not asserted necessary in every special case. Almost sure finiteness alone does not generally supply the required integrability or uniform integrability. These are standard stopping considerations [12, Optional Stopping Theorem, item 4]. For fixed finite homogeneous absorbing chains, the appropriate transient-chain conditions imply finite expected absorption time [11]. At positive error, (18) retains its cooperation-imbalance term in the limit; it is not replaced by (25).

Dividing a bounded stopped total by an ever larger offered horizon can force a zero limit because most of that horizon is never used. It does not recover a continuing payoff ratio. Likewise, at unit slope the noiseless completed-encounter difference can be $-\mu p/\phi$, rather than zero: the label identifies the continuing construction, not a fairness guarantee for stopped totals.

4.4. Collision hazard and complete randomisation

Proposition 4 (execution-error hazard). Under independent execution flips on both players with $\varepsilon > 0$, every intended-action profile has conditional DD probability at least ε^2. Therefore

$$S_H \le (1-\varepsilon^2)^H, \mathbb{E}\tau \le \varepsilon^{-2}. \quad (26)$$

Proof. Conditional on any intended action pair, each player's realised D probability is at least ε. Independent flips give the product lower bound. Averaging over intended actions preserves it, even if those intentions depend on history. Iterating conditional survival bounds proves the first inequality; summing the geometric tail proves the second. ▫

This hazard conclusion needs the joint error assumption, unlike Proposition 3, which needs only X's marginal conditional law. Noise on X alone would not force collision against an opponent that always realises C. Correlated errors require a separate joint-hazard bound.

At $\varepsilon = 1/2$, intended policies no longer affect realised outcomes: each round is independent and uniform over the four states. Direct calculation gives

$$S_H = (3/4)^H, \Pr(\tau = t) = 1/4\,(3/4)^{t-1}, D_H = 4\{1 - (3/4)^H\}, (27)$$

and

$$U_X^{\mathrm{inc}}(H) = U_Y^{\mathrm{inc}}(H) = (T + R + S + P)\{1 - (3/4)^H\}. (28)$$

Equation (18) must not be divided by $1 - 2\varepsilon$ at this endpoint. Persistent positive error implies eventual collision, not any particular finite-horizon ranking.

5. Numerical methods and matched results

5.1. Common specification and exact expectation recursion

All principal comparisons use $(T, R, S, P) = (5,3,1,0)$ and the same fictitious CC seed. The continuing commitments are the slope-two and unit-slope policies from (7) at $\phi = 0.08$. Table 1 states every named policy. Only X's policy changes within a matched comparison.

Role	Intended cooperation vector, in the player's own perspective
X: slope two, $\chi = 2$	$(0.76,0.28,0.24,0)$
X: unit slope, $\chi = 1$	$(1,0.68,0.32,0)$
X and Y reference rules	ALLC: $(1,1,1,1)$; ALLD: $(0,0,0,0)$; TFT: $(1,0,1,0)$
Y: interior rule	$(0.8,0.4,0.6,0.2)$
Y: yield-after-go	$(0,1,0,1)$

Table 1. Intended policies. The names are labels for the displayed vectors; they do not imply stopped-game optimality. "Unit slope" replaces a fairness label because stopped rewards need not be equal. The principal horizon is $H = 50$ and error probability $\varepsilon = 0.02$.

Let $u_0 = \mu$ be alive mass before the first transition. At attempt t, form $w_t = u_{t-1}M$, record first-collision mass $d_t = w_{t,DD}$, then set $u_t = (w_{t,CC}, w_{t,CD}, w_{t,DC}, 0)$. Rewards are added from the full arrival vector w_t, so DD is included once before its mass is removed. Consequently

$$S_H = u_H \mathbf{1}, C_H = \sum_{t=1}^{H} d_t\,, D_H = \sum_{t=1}^{H} u_{t-1}\,\mathbf{1}, U_i^{\mathrm{inc}} = \sum_{t=1}^{H} w_t\, s_i. (29)$$

The survivor tail is right-censored, not a collision assigned to round $H + 1$. For $h \geq 1$, the endpoint law is $(u_{h,CC}, u_{h,CD}, u_{h,DC}, C_h)$; at $h = 0$ it is μ.

For a non-DD seed and fixed policies, let $Q = M_{\{CC,CD,DC\},\{CC,CD,DC\}}$ and $g_i = M_{\{CC,CD,DC\},\cdot}s_i$. Then

$$U_i^{\mathrm{inc}}(H) = \mu_T \sum_{k=0}^{H-1} Q^k\, g_i, D_H = \mu_T \sum_{k=0}^{H-1} Q^k\, \mathbf{1}. (30)$$

If $\rho(Q) < 1$, the infinite sum is $(I - Q)^{-1}$. This is the usual absorbing-chain calculation [10,11], not a separate payoff-control theorem. A fictitious DD seed uses its row for the first transition, followed by the same killed recursion. Encoding it as initially absorbed would instead change the model.

We evaluated five X policies against five fixed Y policies on 103 error values: $0, 0.005, \ldots, 0.5$, with 0.001 and 0.002 added. Each sweep record contains the $H = 50$ stopped outcome and continuing average. Horizon trajectories through 100 rounds were retained at $\varepsilon = 0, 0.02, 0.05, 0.5$. These are probability sums, not Monte Carlo estimates. Appendix B gives independent checks and the reproducibility specification.

5.2. Stationary control and stopped performance

At zero error, the principal slope-two policy against ALLC has stationary payoffs $(4,2)$, attaining Proposition 2. A fixed interior opponent instead produces collisions that can recur in the continuing model. Under first-collision termination, the noiseless infinite-encounter quantities against that interior rule are

X policy	$\mathbb{E}\tau$	U_X^{inc}	U_Y^{inc}	**Controlled residual**	$\rho(Q)$
Slope two	5.35556	14.25556	11.87778	-9.5	0.75740
Unit slope	19.73881	49.96642	62.46642	-12.5	0.94406

Table 2. Noiseless completed encounters from CC against the interior rule. The residual is $U_X - \chi U_Y$ because $P = b = 0$, and equals $-p_{CC}/\phi$. The spectral radii justify finite expected collision time for these examples.

With $\varepsilon = 0.02$ and $H = 50$, survival and total utility differ substantially across the same fixed policies (Table 3). X earns more than Y under the slope-two rule, but its total is 25.5634 below the unit-slope replacement. Within-encounter advantage over the opponent and advantage over another X policy are different comparisons.

X policy	S_{50}	$\mathbb{E}L_{50}$	U_X	U_Y
Slope two	1.1414×10^{-6}	5.2619	13.9279	11.6436
Unit slope	0.031413	15.6643	39.4913	48.6827
ALLC	0.765088	44.0138	109.6059	153.0673
ALLD	7.1422×10^{-19}	2.3805	6.7664	1.5165
TFT	0.005078	11.0567	28.4439	31.9266

Table 3. Stopped outcomes against the interior Y rule, CC seed, $H = 50$, $\varepsilon = 0.02$. Rewards are unconditional accumulated utilities. Included and excluded payoffs coincide here because $P = 0$; very small survival values are not zero.

ALLC has the greatest survival and X total among these five policies against this particular opponent. It also gives Y a larger total. This is not an optimal-policy theorem: the selected policies are not an exhaustive policy space, and no distribution over opponents is assumed. Figure 1 separates survival from first-collision timing and reward accumulation.

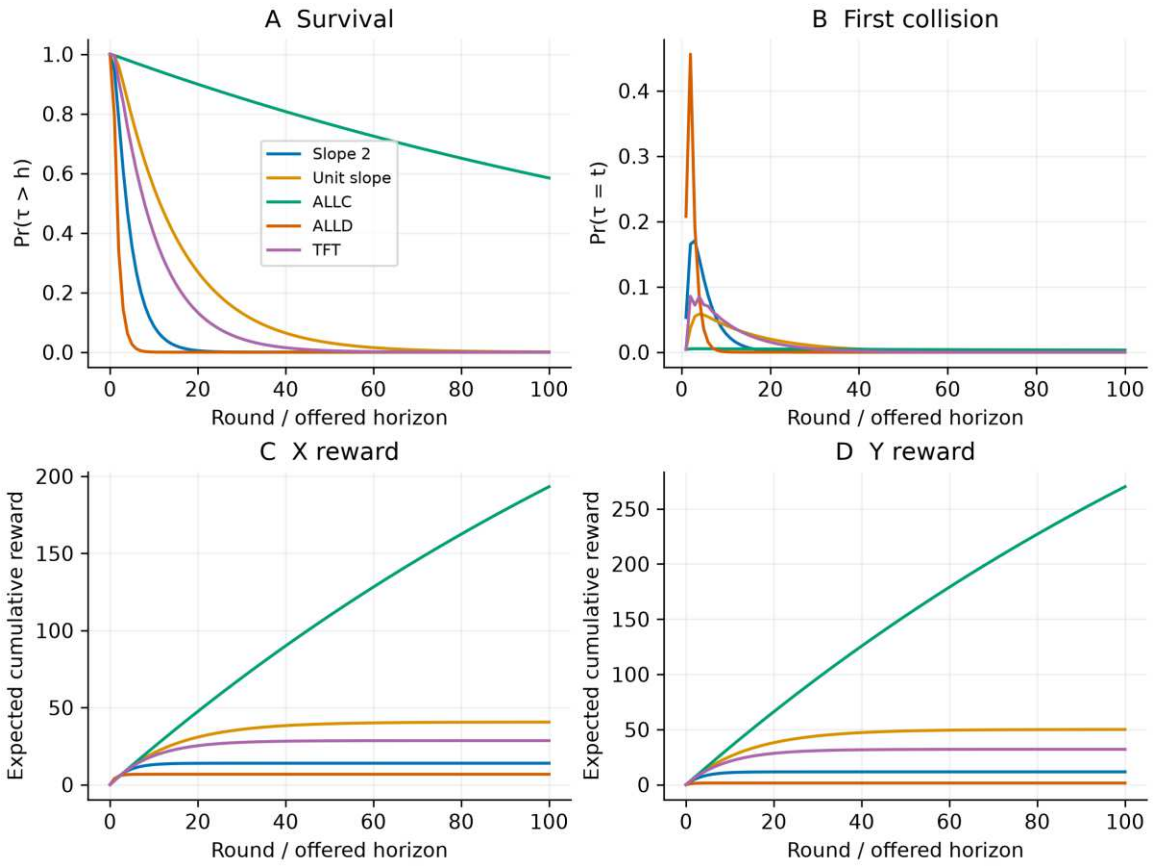


Figure 1. Fictitious CC seed, $\varepsilon = 0.02$, principal payoff tuple. A: probability of no DD through horizon h. B: unconditional first-DD mass at round t. C–D: included expected accumulated utilities. Round 0 is initialisation, not a played round. Curves are deterministic expectations; sampling-error bars do not apply.

5.3. Ranking reversals across payoff criteria, error and horizon

Table 4 compares the slope-two policy with the unit-slope policy against the same Y rule, using a common seed and error rate. Continuing differences are per-round stationary quantities, whereas stopped differences are total encounter utilities. Only their signs are juxtaposed.

Y rule	**Continuing π_X, slope two**	**Continuing π_X, unit slope**	$\Delta\pi_X$	**Stopped** $\Delta U_X(50)$
Interior	1.5814	1.9431	-0.3617	-25.5634
ALLC	3.9565	3.1061	$+0.8505$	$+6.0300$
ALLD	0.1318	0.1708	-0.0389	-2.0780
TFT	0.3670	1.5834	-1.2164	-80.6938
Yield-after-go	3.3606	2.6130	$+0.7477$	-2.0210

Table 4. Slope-two minus unit-slope comparisons at $\varepsilon = 0.02$, CC seed, principal tuple. Stationary enforcement in Proposition 1 is a zero-error result; these positive-error stationary values are not asserted to lie on the intended noiseless payoff line.

Against yield-after-go, the continuing advantage of 0.7477 becomes a stopped disadvantage of 2.0210. The stopped X totals are 1.5726 and 3.5936. This example disproves a general transfer of the same-policy stationary ranking to the stopped objective under these specified conditions. It does not show that either policy is best in the stopped game.

Error can also reverse a stopped ranking against a single opponent. Against ALLC with no error and $H = 50$, slope two yields X total 198.9167, compared with 150 at unit slope. At $\varepsilon = 0.02$, the totals are 155.7868 and 149.7568, while survival is approximately 0.6085 and 0.9335. At $\varepsilon = 0.05$, the totals are 110.8600 and 133.9245, favouring unit slope. A bracketed numerical root in $[0.02,0.05]$ occurs at $\varepsilon = 0.02408096$. This does not establish that no other crossover exists.

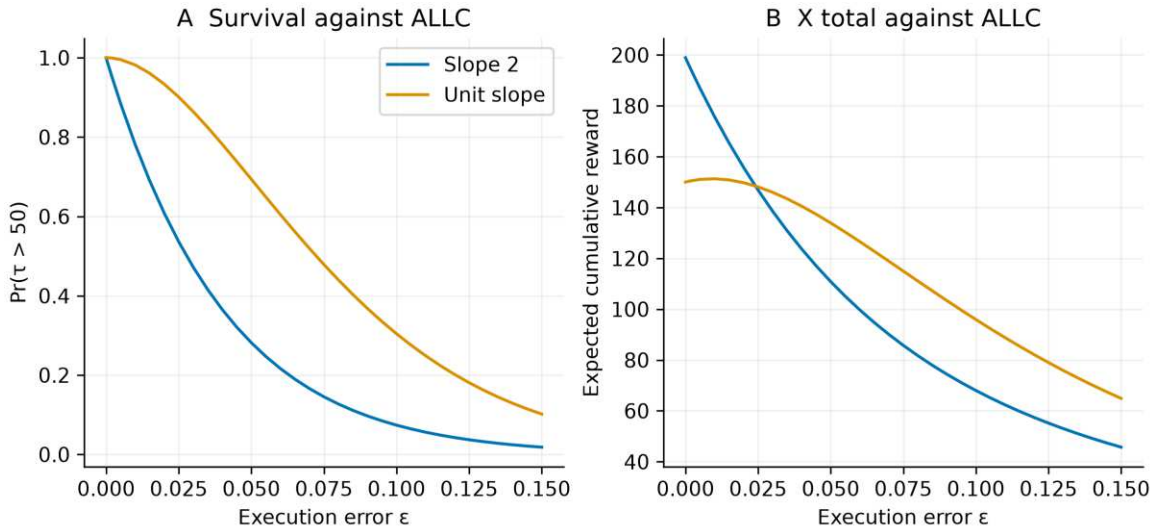


Figure 2. ALLC opponent, CC seed, $H = 50$, principal tuple. A: survival. B: X accumulated utility. Intended policies are held fixed. The crossing in B is near $\varepsilon = 0.02408$; it is not an equality of the survival curves.

Keeping ALLC and $\varepsilon = 0.02$ fixed while extending the offered horizon also changes the stopped comparison: the slope-two minus unit-slope difference is $+0.1923$ at $H = 60$ and -0.4940 at $H = 61$. Figure 3 compares continuing and stopped differences over the error range. No benefit attributed to error in the stopped model can involve recovery after a collision, because no such recovery is allowed.

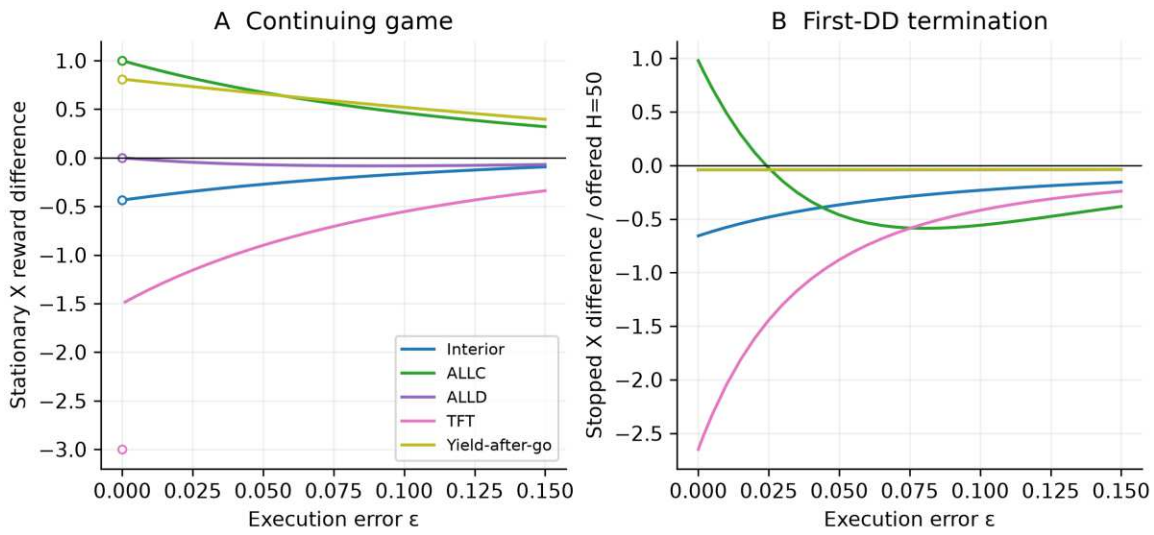


Figure 3. Slope-two minus unit-slope X rewards against each named Y rule. A: continuing stationary difference; open circles are the separate seed-selected zero-error Cesàro values. B: stopped cumulative difference divided by the fixed offered $H = 50$. A positive value favours slope two within that panel's criterion. This division preserves the stopped comparison's sign but does not equate the two criteria.

5.4. Initialisation and terminal accounting

The yield-after-go comparison depends on the opening-memory law (Table 5). A common fictitious DC seed gives slope two the larger stopped total, reversing the CC-seed result. A uniform memory law also gives a small positive difference. Against the interior opponent, unit slope gives the higher X total under every tested seed; the full values are in the accompanying records.

Fictitious seed against yield-after-go	Slope-two U_X	Unit-slope U_X	Difference
CC	1.5726	3.5936	-2.0210
CD	0.8899	2.5985	-1.7085
DC	18.1037	15.9532	$+2.1506$
Uniform	10.7114	10.5998	$+0.1115$

Table 5. Same policies, $\varepsilon = 0.02$, $H = 50$ and payoff tuple. Uniform assigns equal mass to all four fictitious memories, including DD; its DD mass does not terminate before round 1. A common seed law does not imply the same opening actions across X policies.

A separate terminal-charge sensitivity changes the collision reward from 0 to -2 while holding all intended policies and transition probabilities fixed. Against the interior opponent, the slope-two X total falls from 13.9279 to 11.9279, whereas ALLC falls from 109.6059 to 109.1361. Each change is $-2C_{50}$, as required by (6). Survival and excluded rewards are unchanged. These fixed policies are not asserted to remain baseline-P ZD policies for the altered reward matrix. Reconstructing the policy after changing P would instead alter behaviour and would answer a different question.

The nonzero-baseline example (24) is a different diagnostic: it reconstructs an admissible policy for $P = b = -1$ and checks the two affine-exclusion conventions. Both tests are needed to avoid conflating changed utility accounting with changed strategic commitment.

6. Discussion

The stationary construction and stopped encounter have a common cooperation-flow basis, but their objectives are distinct. Stationarity removes the net flow between consecutive rounds. First-collision stopping retains a systematically selected final state, so the endpoint generally remains. At zero error and finite expected collision time it can be a nonzero constant even as the offered horizon grows. Execution error adds a second term; it should not be used to explain a discrepancy already present without error.

A cooperative best response for stationary payoffs is compatible with lower stopped rewards under that response criterion or another initialisation. Proposition 2 has explicit stage-payoff and zero-error assumptions. It is neither a theorem about the stopped optimum nor a statement about adjustment dynamics. The scale-derivative calculation in Appendix A is also not an opponent best-response calculation: varying X's commitment scale differs from changing Y's policy.

Collision exclusion is an accounting choice rather than deletion of a boundary state. Removing a terminal payoff does not remove its potential from a telescoped identity. If the duration counts only safe rounds, its baseline term must change as well. An absorbing-game encoding should therefore specify the reward on entry to the cemetery and assign zero rewards after entry. A persistent absorbing state paying a nonzero amount forever would define a different total-reward problem. A fictitious initial DD memory must also remain distinct from actual absorption.

The numerical comparisons have limited scope. They use one main payoff tuple, one scale and a named set of opponent rules; no empirical utility, error rate or initialisation law is estimated. Symmetric independent execution errors exclude asymmetric implementation and correlated shocks. The examples provide counterexamples to automatic ranking transfer, not a ranking valid across all Chicken games or all opponents. We do not classify all stopped-game payoff-controlling policies, and the corrected identity is not an impossibility result for other constructions or reward coordinates.

The established finite-time and multi-state literature is part of the derivation, not a competing model to be dismissed. The application here is explicit first-DD accounting with a fixed continuing-game family, a stated execution-error model and reproducible comparisons. Exact priority of this combination is not asserted. The question of optimising stopped policies subject to a survival constraint requires a specified objective and a separate analysis.

7. Conclusion

A single cooperation-flow construction connects stationary payoff control to collision-terminated Chicken, provided its endpoint is retained. The admissibility constraints and an attainable stationary opponent bound can be stated independently of any claim about learning. At first collision, cumulative rewards instead satisfy an endpoint-corrected identity, with an additional execution-error correction and distinct terminal-exclusion conventions. Matched calculations

show that continuing and stopped criteria can rank the same commitments differently, while initialisation can reverse the stopped comparison. Payoff control and survival should therefore be evaluated under the actual continuation and reward rules rather than inferred from a stationary payoff line.

Although the paper is written entirely in terms of the repeated Chicken game, its broader contribution is methodological. It argues that models of strategic behaviour should reflect the actual operational rules of the system being analysed. If the real system can end irreversibly after a failure, then conclusions drawn from stationary, infinite-horizon averages may be misleading. For safety-critical systems, the appropriate objective is often performance before failure, not performance averaged over an interaction that is assumed never to end. This insight extends beyond game theory to any domain where catastrophic events terminate future opportunities, making the stopping rule itself part of the strategic problem rather than a technical detail. This interpretation is consistent with the paper's conclusion that payoff control and survival must be evaluated under the actual continuation and reward rules of the interaction, rather than inferred from stationary payoff relationships.

Appendix A. Determinant reduction and scale response

The stationary balance proof does not require determinants. A transition-consistent determinant nevertheless clarifies scale sensitivity. This appendix is restricted to the noiseless continuing model and baseline P. Write $h = f_P$ and $p = e + \phi h$. For any outcome vector z, define

$$D(z) = \det[m_1, p - e, r - (1,0,1,0)^T, z], (A1)$$

Here m_1 is the first column of $M - I$. The first three displayed columns are obtained by retaining that first column and adding it to the original second and third columns. Their stationary left null vector is v. Where $D(\mathbf{1}) \neq 0$, cofactor normalisation gives $vz = D(z)/D(\mathbf{1})$. If the denominator vanishes, recurrent-class analysis is required; Proposition 1 remains valid without this quotient.

Set $u = e \odot r - (1,0,0,0)^T$ and $d = r - (1,0,1,0)^T$, where $\odot$ denotes componentwise multiplication. Define

$$A_z = \det[u, h, d, z], B_z = \det[h \odot r, h, d, z]. (A2)$$

Multilinearity gives the exact factorisation

$$D(z) = \phi(A_z + \phi B_z), \pi_Y = \frac{A_{s_Y} + \phi B_{s_Y}}{A_{\mathbf{1}} + \phi B_{\mathbf{1}}}. (A3)$$

Therefore, holding q and χ fixed where the quotient is defined,

$$\frac{d\pi_Y}{d\phi} = \frac{B_{s_Y} A_{\mathbf{1}} - A_{s_Y} B_{\mathbf{1}}}{(A_{\mathbf{1}} + \phi B_{\mathbf{1}})^2}. (A4)$$

The derivative sign is constant in scale on such an interval, but need not be independent of χ. For the principal tuple, $q = (0.28,0.74,0.88,0.58)$ and $\phi = 0.03$, admissible choices give $\pi_Y = 2.31765125$ and derivative $+0.30281623$ at $\chi = 1.2$, but $\pi_Y = 0.88391988$ and derivative -0.10865473 at $\chi = 3.8$. Both are continuing-game calculations. The sign of this derivative is not a statement about how Y changes its own conditional probabilities, and no stability conclusion follows without a specified adjustment rule.

Appendix B. Reproduction and independent numerical checks

The accompanying standalone script `reproduce.py` uses NumPy, SciPy and Matplotlib. It imports no application code or earlier manuscript calculations. Run it in its containing directory, or pass its path to Python from another directory; outputs are written beside the script. It regenerates all numerical records and the three figure files. The archive includes full-precision JSON results and CSV versions of the main, comparison, initialisation and infinite-encounter tables.

For every named policy pair, an independent realised-history tree explicitly sums intended actions and flip events for horizons 1–4, seeds CC, uniform and fictitious DD, errors $0,0.02,0.17,0.5$, and collision rewards $0,-2$. Its 2,400 horizon-level comparisons do not use the transition-matrix reward recursion. Additional history-tree checks use an adaptive Y whose conditional action depends on the parity of its earlier realised cooperation count; these test the stopped identity outside the four-state opponent model. Such examples check the implementation, not the general theorem.

Check	Comparisons	Maximum absolute error
Independent short-history enumeration	2,400	1.59×10^{-13}
Probability conservation over all sweep horizons	260,075	6.00×10^{-15}
Noiseless stopped identity	1,010	5.82×10^{-14}
Noisy stopped identity, $\varepsilon < 1/2$	102,010	3.27×10^{-12}
Adaptive-opponent stopped identity	30	2.14×10^{-14}
Complete-randomisation formulas	100	5.33×10^{-15}
Sampled stationary enforcement	1,000	1.34×10^{-14}
ALLC attainment, sampled interior parameters	1,000	1.78×10^{-15}
Determinant quotient	1,003	3.78×10^{-15}
Scale derivative versus differentiated stationary equations	1,003	1.12×10^{-12}

Table B1. Errors are conservatively rounded upward. All checks passed an absolute tolerance of 10^{-9}. The archive also records upper-bound checks, slope-endpoint attainment, infinite-encounter residuals and nonzero-baseline accounting. There are 370,637 scalar check records in total; the counts include repeated pointwise consistency checks, not independent observations.

Stationary validation uses 1,000 parameter sets with pseudorandom seed 20261109: χ uniform on $(1.05,4.95)$, the scale a uniform $(0.05,0.95)$ fraction of $1/(5\chi-1)$, and each opponent component uniform on $(0.02,0.98)$. Three named scale-response cases supplement them. The derivative check differentiates $vM = v$, $v\mathbf{1} = 1$ and compares with (A4), rather than using a finite-difference approximation. Slope endpoints $\chi = 1,5$ are checked separately against ALLC.

The ALLC error crossover uses Brent's method within $[0.02,0.05]$, with error-rate tolerance 10^{-14} and final payoff-difference residual below 5×10^{-13}. No claim is made about crossings outside that bracket. These deterministic calculations have no sampling uncertainty to estimate

by confidence intervals. Their relevance depends on the specified utilities and interaction assumptions.

Data and code availability

No empirical observations were used. The accompanying reproducibility archive contains the standalone calculation script, parameter specification, validation records, full-precision numerical results and figure files. It is sufficient to regenerate the reported expectations without the earlier manuscripts or an application installation. A public archival identifier is not asserted.

References

[1] Press, W. H., & Dyson, F. J. (2012). Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences*, 109(26), 10409–10413. https://doi.org/10.1073/pnas.1206569109

[2] Akin, E. (2015). What You Gotta Know to Play Good in the Iterated Prisoner's Dilemma. *Games*, 6, 175–190. https://doi.org/10.3390/g6030175

[3] Ueda, M. (2026). Performance of Zero-determinant Strategies in Repeated Games without Discounting. *arXiv preprint*, 2610.02645v1. https://arxiv.org/abs/2610.02645v1

[4] McAvoy, A., & Hauert, C. (2016). Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences*, 113, 3573–3578. https://doi.org/10.1073/pnas.1520163113

[5] Ichinose, G., & Masuda, N. (2018). Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology*, 438, 61–77. https://doi.org/10.1016/j.jtbi.2017.11.002

[6] Govaert, A., & Cao, M. (2020). Zero-Determinant strategies in repeated multiplayer social dilemmas with discounted payoffs. *arXiv preprint*, 1910.07858v2. https://arxiv.org/abs/1910.07858v2

[7] Palasciano, M. (2020). Autocratic Strategies of Multi-State Games. *arXiv preprint*, 2012.14566v1. https://arxiv.org/abs/2012.14566v1

[8] McAvoy, A., Madhushani Sehwag, U., Hilbe, C., Chatterjee, K., Barfuss, W., Su, Q., Leonard, N. E., & Plotkin, J. B. (2025). Unilateral incentive alignment in two-agent stochastic games. *Proceedings of the National Academy of Sciences*, 122, e2319927121. https://doi.org/10.1073/pnas.2319927121

[9] Mamiya, A., Miyagawa, D., & Ichinose, G. (2021). Conditions for the existence of zero-determinant strategies under observation errors in repeated games. *arXiv preprint*, 2105.03951v2. https://arxiv.org/abs/2105.03951v2

[10] Tan, L., Mahdaviani, K., & Khisti, A. (2021). Markov Rewards Processes with Impulse Rewards and Absorbing States. *arXiv preprint*, 2105.00330v1. https://arxiv.org/abs/2105.00330v1

[11] Grinstead, C. M., & Snell, J. L. *Introduction to Probability*, Chapter 11, Section 11.2, Theorems 11.3–11.5. Dartmouth College electronic text. https://math.dartmouth.edu/~prob/prob/prob.pdf

[12] Wu, H. (2015). *MIT 18.445: Introduction to Stochastic Processes*, Lecture 16, 13 April, Optional Stopping Theorem. MIT OpenCourseWare. https://ocw.mit.edu/courses/18-445-introduction-to-stochastic-processes-spring-2015/ddeda684b0dd6a1422418c7eca3ce843_MIT18_445S15_lecture16.pdf