Temporal Fair Division in Multi-Agent Systems: From Precise Alternation Metrics to Scalable Coordination Proxies

NIKOLAOS AL. PAPADOPOULOS*, University of Macedonia, Greece

The fair allocation of shared resources is among the most studied problems in economics and computer science. While classical fair division addresses static, one-shot assignments, many real-world settings require agents to compete for the same resource repeatedly over time, calling for a *temporal* notion of fairness in which equitable access is judged across an entire history of interactions rather than at any single moment. This paper contributes to the emerging theory of temporal fair division by introducing and analysing a lightweight metric, Rotational Periodicity (RP), alongside a family of precise sliding-window measures (ALT metrics), within a unified evaluation framework for repeated multi-agent resource competition.

We formalise the Multi-Agent Battle of the Exes (MBoE) as a repeated fair division instance and establish Perfect Alternation (PA) as its canonical temporally fair solution, drawing explicit connections to proportionality, envy-freeness, and n-periodic round-robin allocation. RP decomposes temporal fairness into two complementary sub-measures (Rotational Score, RS, and Waiting Periods Evaluation, WPE) achieving linear time complexity $O(v + n)$ versus the quadratic $O(v \cdot n)$ of ALT, where v is the number of episodes and n is the agent count.

Empirical evaluation across agent populations of $n \in \{2, 3, 5, 8, 10\}$ reveals three key findings. First, both RP and ALT expose a *coordination failure* undetectable by traditional metrics for $n \geq 3$: Q-learning agents perform *worse* than random policies by 10–73% on RP and 7–35% on CALT, with the gap peaking at $n = 3$ and narrowing as n grows (since random policies themselves approach the PA ideal by the law of large numbers), while Reward Fairness remains misleadingly high (> 0.92 for $n \geq 3$). Second, RP achieves a computational speedup of 12–25× over the ALT family, with the advantage growing with agent count. Third, the two families capture complementary aspects of temporal fairness: ALT provides richer discrimination for small populations whereas RP scales reliably to larger systems where ALT becomes intractable. Together, they form a diagnostic toolkit for assessing coordination quality in temporal fair division settings.

CCS Concepts: • **Computing methodologies → Multi-agent reinforcement learning**; • **Theory of computation → Algorithmic game theory and mechanism design**; *Online algorithms and data structures.*

Additional Key Words and Phrases: temporal fair division, turn-taking, multi-agent systems, rotational periodicity, alternation metrics, coordination, reinforcement learning, resource allocation

ACM Reference Format:
Nikolaos Al. Papadopoulos. 2026. Temporal Fair Division in Multi-Agent Systems: From Precise Alternation Metrics to Scalable Coordination Proxies. *ACM Trans. Econ. Comput.* 0, 0, Article 0 (2026), 16 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Corresponding author.

Author's Contact Information: Nikolaos Al. Papadopoulos, nikolaos.papadopoulos@uom.edu.gr, University of Macedonia, Department of Applied Informatics, Thessaloniki, Greece.

ACM 2167-8383/2026/0-ART0
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The fair division of goods and resources is a foundational concern in both economic theory and algorithmic game theory [2, 12]. The vast majority of this literature treats allocation as a one-shot problem: given a set of items and a set of agents with preferences, find an assignment satisfying some combination of fairness and efficiency axioms such as envy-freeness [5], proportionality [20], or maximum Nash welfare [3].

Yet many environments of practical significance are *repeated*: agents compete for the same limited resource episode after episode, and the question is not merely who obtains the resource today, but whether every agent obtains it *often enough and regularly enough* over the long run. Network bandwidth allocation, smart-grid energy distribution, traffic signal control, and turn-based collaborative robotics all share this structure [8, 11]. In these settings the relevant fairness criterion is *temporal*: an allocation scheme is temporally fair if the sequence of wins for each agent, considered as a whole, satisfies some notion of balance and regularity.

This temporal dimension has received growing attention under several labels: online fair division [7], repeated allocation [1], and sequential resource sharing; yet the *measurement* problem is underexplored. How do we quantify how close a system is to the temporal fairness ideal? How do we compare agents' histories in a way that is both sensitive to coordination failures and computationally tractable for large populations?

The present paper addresses these questions by studying temporal fair division in the Multi-Agent Battle of the Exes (MBoE), a repeated competitive game in which n self-interested agents repeatedly attempt to claim a shared high-reward resource. We make three contributions.

Contribution 1: Framework. We formalise MBoE as a repeated fair division instance and identify *Perfect Alternation* (PA), a Pareto-optimal coordination regime in which every agent wins exactly once in every n-episode block, as the canonical temporally fair solution (a Nash equilibrium in the two-agent ballistic formulation [15]). We show that PA satisfies temporal proportionality and a form of temporal envy-freeness analogous to the classic EF notion in static fair division.

Contribution 2: Rotational Periodicity. Building on a conference precursor [14], we formalise and empirically validate Rotational Periodicity (RP), a family of lightweight metrics that decompose temporal fairness into two orthogonal dimensions: *Rotational Score* (RS), a symmetric ratio measuring how close each agent's mean inter-win gap is to the ideal $n - 1$ episodes—replacing the asymmetric AWE sub-measure of the conference precursor [14]—and *Waiting Periods Evaluation* (WPE), measuring whether the *count* of gaps approximates the ideal ν/n. Each sub-measure has two event-type variants (exclusive wins, all terminal reaches) yielding six sub-metrics ($\mathrm{RS}_{\mathrm{excl}}$, $\mathrm{RS}_{\mathrm{reach}}$, $\mathrm{WPE}_{\mathrm{excl}}$, $\mathrm{WPE}_{\mathrm{reach}}$, and the deprecated AWE variants); combinations of these form four named $\overline{\mathrm{RP}}$ variants (Section 4.3). All variants are normalised against the PA baseline and computed in $O(\nu + n)$ time. Weighted (FRP) and equitable (ERP) extensions support heterogeneous-priority settings.

Contribution 3: Comparative Empirical Study. We present the first systematic comparison of RP with the ALT family of sliding-window metrics [13, 15] across agent populations $n \in \{2, 3, 5, 8, 10\}$, using both Q-learning policies and random baselines. The comparison spans metric expressiveness, sensitivity, and computational cost, and leads to concrete guidance on when each family is preferable.

Our experiments uncover a striking result: for $n \geq 3$, Q-learning agents achieve high Reward Fairness (> 0.92) yet score *worse than random policies* on both RP and CALT (Comprehensive ALT) by margins of 7–73%, with the gap peaking at $n = 3$ (where random already alternates well by chance) and narrowing as the random baseline itself approaches the PA ideal at large n. For $n = 2$, Q-learning converges to a monopoly strategy (one agent always wins), yielding *low* Reward Fairness (0.49) alongside low RP and CALT: traditional metrics also detect failure here, making

$n = 2$ a distinct regime. This *coordination gap* [15] is invisible to traditional metrics and underscores the necessity of temporal fairness measures for faithful evaluation of multi-agent coordination.

The remainder of the paper is organised as follows. Section 2 surveys related work. Section 3 formalises MBoE and its connections to fair division theory. Section 4 defines all metrics in the framework. Section 5 provides a formal analytical comparison. Section 6 presents the empirical study. Section 7 discusses implications. Section 8 concludes.

2 Related Work

2.1 Static Fair Division

The theoretical study of fair allocation traces back to the "problem of fair division" posed by Steinhaus [20] and formalised through the notions of proportionality and envy-freeness [5]. For divisible goods ("cake-cutting") a rich literature establishes existence and complexity results [2]. For indivisible goods (closer to our setting), Lipton et al. [10] introduced Envy-Freeness up to one Good (EF1), showing it is always achievable, while subsequent work established stronger notions and their computational properties [3]. The textbook by Moulin [12] provides a comprehensive treatment of fair division from a welfare economics perspective.

2.2 Sequential and Repeated Allocation

When items must be allocated one at a time over multiple rounds, round-robin mechanisms are natural candidates. Bouveret and Lang [1] study the fairness properties of such protocols for indivisible goods, showing that n-periodic round-robin achieves proportionality in an approximate sense. He et al. [7] investigate fairness over time in an online setting where items arrive sequentially and allocations must be made irrevocably. Our setting differs in that the same resource is competed for in every round, so the allocation problem is genuinely repeated rather than sequential.

2.3 Turn-Taking in Multi-Agent Systems

Turn-taking as an emergent coordination behaviour has been studied in both biological and artificial agent settings [4, 18]. The Battle of the Exes (BoE), introduced by Hawkins and Goldstone [6], provides an idealised game in which two agents must learn to alternate access to a high-reward location without communication. Papadopoulos and Sanchez-Fibla [13, 15] extended this game to n agents (MBoE) and proposed the ALT family of metrics for evaluating turn-taking quality. Raffensperger et al. [17] proposed a simpler metric for emergent turn-taking in communication experiments. RP was first proposed as a scalable complement to ALT in a conference paper [14]; the present work extends that preliminary proposal with a full formalisation, a fair division framing, and the first systematic empirical comparison of RP and ALT across $n \in \{2, 3, 5, 8, 10\}$. In a companion paper, Papadopoulos and Psannis [15] conduct a large-scale study of coordination failure in MBoE, showing that Q-learning policies consistently underperform random baselines on ALT metrics across all tested configurations. The present paper complements that study by introducing RP, performing the first direct comparison of RP and ALT, and embedding both within a fair division framework.

2.4 Multi-Agent Reinforcement Learning and Common-Pool Resources

Multi-agent reinforcement learning (MARL) has been applied extensively to resource sharing and common-pool problems [9, 16]. A persistent challenge is that standard reward signals do not incentivise coordination beyond what is captured by cumulative reward. Perolat et al. [16] show that agents in common-pool resource games can learn behaviours that deplete the resource despite achieving high individual returns. Our findings echo this: high Reward Fairness coexists

with coordination failure, confirming that temporal metrics are necessary additions to the MARL evaluation toolkit.

3 Multi-Agent Battle of the Exes as Temporal Fair Division

3.1 Game Formulation

The Multi-Agent Battle of the Exes (MBoE) with n agents proceeds in discrete episodes $t = 1, 2, \ldots, \nu$. At each episode t, every agent $i \in \{1, \ldots, n\}$ independently attempts to reach a terminal position. The payoff structure is:

- A *solo winner* (the unique agent reaching its terminal position) receives r_{high}.
- Under *Inverse Linear Fractional* (ILF) rewards, agents in a tied arrival receive r_{high}/n.
- Under *Inverse Quadratic Fractional* (IQF) rewards, agents in a tied arrival receive r_{high}/n^2.
- Agents that do not reach their terminal position receive zero.

We set $r_{\text{high}} = 100$ throughout. Agents observe the game state and update a Q-table according to the standard tabular Q-learning rule with learning rate $\alpha = 0.3$, discount factor $\gamma = 0.999$, and an ε-greedy policy with ε decaying linearly from 0.9 to 0.004 over 75% of the episode budget.

Connection to fair division. At each episode the right to be the sole winner (the "high-value resource") can be allocated to at most one agent. Over ν episodes the resource is available ν times; perfect efficiency requires that it is claimed every episode, while perfect fairness requires that each agent claims it the same number of times. This is precisely a repeated fair division problem in which the item of value is the exclusive access right, and the allocation is determined endogenously by the agents' strategies.

3.2 Two State Representations

Type-A states encode only agent positions: $s_t = [p_1^t, \ldots, p_n^t]$. **Type-B** states additionally include a memory vector recording, for each agent i, whether it was the sole winner in the previous episode: $s_t = [p_1^t, \ldots, p_n^t, z_1^t, \ldots, z_n^t]$, where $z_i^t \in \{0, 1\}$. Type-A serves as the primary experimental condition; Type-B is included for completeness.

3.3 Perfect Alternation as the Temporal Fairness Ideal

Definition 3.1 (Perfect Alternation Equilibrium). An outcome is said to satisfy *Perfect Alternation* (PA) if, over every block of n consecutive episodes, each agent is the sole winner exactly once. Equivalently, the sequence of winners forms an n-periodic cycle in which every agent appears exactly once per period.

PA is Pareto-optimal and constitutes a Nash equilibrium in the two-agent ballistic formulation [15]; for $n > 2$ it serves as the canonical temporally fair reference point. PA satisfies the following temporal analogues of classical fair division criteria:

Temporal Proportionality. Each agent wins ν/n times in ν episodes, receiving exactly $1/n$ of the total high-value resource.

Temporal Envy-Freeness. No agent i strictly prefers the win-history of any other agent j to its own: since all agents win at the same rate with the same inter-win gaps, no temporal envy arises.

Regularity. The inter-win gaps for each agent are identically equal to $n - 1$, minimising uncertainty about resource access timing.

The PA concept is analogous to n-periodic round-robin allocation studied in the sequential allocation literature [1]: each agent takes exactly one turn per round of n slots. The key difference

is that in MBoE, the alternating schedule must *emerge* from independent agent learning rather than being enforced by a central planner.

4 Metrics for Temporal Fairness

We present the full set of metrics in the framework, progressing from the coarsest to the finest-grained.

4.1 Traditional Metrics

Let R_i^v denote agent i's cumulative reward through episode v, and let $R_{\max} = \max_i R_i^v$.

Definition 4.1 (Efficiency).

$$E = \frac{\sum_{i=1}^{n} R_i^v}{v \cdot r_{\text{high}}}.$$

E reaches 1.0 when no episode is wasted (solo winner every episode, no ties with shared rewards). Under PA, $E = 1.0$.

Definition 4.2 (Reward Fairness).

$$\text{RF} = \frac{\sum_{i=1}^{n} R_i^v}{n \cdot R_{\max}}.$$

RF approaches 1.0 when rewards are evenly distributed. Under PA, RF = 1.0. As we demonstrate empirically, high RF is a necessary but far from sufficient condition for temporal fairness: random policies can achieve RF > 0.9 without any coordination whatsoever.

4.2 ALT Metrics: A Sliding-Window Family

The ALT family [13, 15] evaluates turn-taking quality by sliding a window of width n across the episode sequence and averaging a batch score β_j over all $v - n + 1$ windows. Formally:

$$\text{ALT} = \frac{1}{v - n + 1} \sum_{j=0}^{v-n} \beta_j.$$

The six variants differ in their definition of β_j. Let batch j span episodes $[j, j + n - 1]$. Define: f_j: number of distinct agents winning at least once in batch j; t_j: total terminal arrivals in batch j; w_j: number of episodes with exactly one winner in batch j; g_j: number of agents achieving at least one *solo* win in batch j; y_k: number of agents winning episode k of batch j.

Table 1. ALT Metric Variants and Their Batch Score Formulas

Metric	β_j formula	Interpretation
Primary metrics (CALT, EALT, AALT)		
CALT	$\frac{\sum_{k=1}^{n}(n - y_k)}{n(n-1)}$	Penalises simultaneous arrivals; most comprehensive single-number summary
EALT	w_j/n	Fraction of episodes with a unique winner
AALT	g_j/n	Fraction of agents achieving at least one solo win per window
Auxiliary variants (additional nuance near PA)		
FALT	f_j/t_j	Fraction of wins accruing to distinct agents
qFALT	$(f_j/n)^2$	FALT with quadratic emphasis near PA
qEALT	$(w_j/n)^2$	EALT with quadratic emphasis near PA

Every ALT metric equals 1.0 under PA and 0 under complete coordination failure. The three primary metrics (CALT, EALT, AALT) capture complementary aspects of coordination quality at different strictness levels: CALT penalises simultaneous arrivals most aggressively; EALT measures exclusivity at episode level; AALT measures per-window agent coverage. The auxiliary variants (FALT, qFALT, qEALT) provide additional discrimination near-perfect alternation and are not the focus of this paper. All ALT computation has complexity $O(v \cdot n)$.

AltRatio and PA-Equivalent Agents. Following [15], the *AltRatio* is x/n, the proportion of agents behaving as if in perfect alternation. For CALT specifically, the quadratic construction of its batch score means the metric scales approximately as $(x/n)^2$ (with a near-zero intercept $\varepsilon \approx 0$ [15]), so:

$$\text{AltRatio}_{\text{CALT}} \approx \sqrt{\text{CALT}}, \quad \textit{PA-equivalent agents} \approx n\sqrt{\text{CALT}}.$$

For EALT and AALT the scoring is linear in x/n, so $\text{AltRatio}_{\text{EALT}} = \text{EALT}$ and $\text{AltRatio}_{\text{AALT}} = \text{AALT}$ directly (no square-root transformation).

4.3 Rotational Periodicity: A Linear-Time Alternative

RP measures temporal fairness by analysing each agent's *waiting pattern* individually (specifically, the sequence of gap lengths between consecutive wins) rather than by scanning the global episode sequence in batches.

Rotational Score (RS).. For agent i with at least two solo wins, let k_i denote the solo-win count, $\bar{r}_i = \frac{1}{k_i - 1} \sum_{j=1}^{k_i - 1} r_{i,j}$ the mean inter-win gap (episodes between consecutive sole victories, exclusive of win episodes), and $r_i^* = n - 1$ the ideal gap under Perfect Alternation. The *Rotational Score* is

$$\text{RS}_i = \begin{cases} \dfrac{\min(\bar{r}_i,\, r_i^*)}{\max(\bar{r}_i,\, r_i^*)} & \text{if } k_i \geq 2, \\ 0 & \text{otherwise.} \end{cases}$$

RS replaces the asymmetric AWE sub-measure used in the conference precursor [14], which applied a hard threshold at $\bar{r}_i = 2r_i^*$ (returning 0 for any gap beyond twice the ideal) and gave the same penalty to an agent waiting $2r_i^*$ as to one waiting $100\, r_i^*$. RS penalises deviations *symmetrically*: an agent winning at twice the ideal rate ($\bar{r}_i = r_i^*/2$) and an agent winning at half the ideal rate ($\bar{r}_i = 2r_i^*$) both receive $\text{RS}_i = 0.5$. Under PA, $\bar{r}_i = r_i^* = n - 1$ and $\text{RS}_i = 1$. As $\bar{r}_i$ diverges from r_i^* in either direction RS_i decreases continuously towards 0, providing a non-trivial signal even for very infrequent winners.

Weighted Temporal Fair Division. In heterogeneous-priority settings where agent i holds target share $w_i > 0$ ($\sum_{i=1}^{n} w_i = 1$), the ideal gap generalises to $r_i^* = 1/w_i - 1$ (the expected inter-win gap when the agent wins w_i of all episodes). The resulting *weighted RS*, $\text{RS}_i^w = \min(\bar{r}_i, r_i^*)/\max(\bar{r}_i, r_i^*)$, equals 1 when the agent's empirical win rate matches its priority w_i exactly. The uniform case ($w_i = 1/n$, $r_i^* = n - 1$) is recovered as a special case. This formulation aligns RS directly with *weighted proportionality*—the standard fairness criterion for agents with heterogeneous claims [12]: each agent's share of the contested resource should equal w_i.

Waiting Periods Evaluation (WPE).. WPE measures whether the *count* of waiting periods is consistent with PA. Under PA, each agent should experience $t_i^* = v/n$ waiting periods (including boundary periods before the first win and after the last). Then:

$$\text{WPE}_i = \begin{cases} 1 - \dfrac{|t_i - t_i^*|}{t_i^*} & \text{if } t_i < 2t_i^*, \\ 0 & \text{otherwise.} \end{cases}$$

WPE captures the *distribution* of access opportunities: whether the agent receives the fair share of turns in terms of frequency.

Combined RP.. The per-agent RP is a weighted combination:

$$\mathrm{RP}_i = \frac{\alpha \cdot \mathrm{RS}_i + \beta \cdot \mathrm{WPE}_i}{\alpha + \beta},$$

and the system-wide metric is $\overline{\mathrm{RP}} = (1/n) \sum_i \mathrm{RP}_i$. In all experiments we set $\alpha = \beta = 1$, giving equal weight to rhythm and frequency. Domain-specific settings can emphasise one component; for example, time-sensitive applications may prefer $\alpha > \beta$.

Win-Event Definition. Both WPE and the AWE sub-formula of the conference precursor [14] require specifying what counts as a *win event* for agent i: either (i) *exclusive wins* (sole victories, episodes in which exactly one agent reaches its terminal position), or (ii) *all terminal reaches* (including simultaneous arrivals). The two definitions produce systematically different behaviour. Exclusive wins are sparse for $n \geq 3$: inter-win gaps routinely exceed $2r_i^*$, collapsing the AWE formula to 0 for all Q-learning agents in our experiments (RS is immune to this collapse because its symmetric ratio min/max has no hard cutoff). All-reaches events, by contrast, occur so frequently that $t_i \gg 2t_i^*$, collapsing $\mathrm{WPE}_i = 0$ for $n \geq 3$. The choice therefore depends on what the researcher wishes to quantify: sole-victory frequency (exclusive wins, measuring individual dominance) or arrival rhythm relative to all interactions (all reaches, appropriate when ties carry information). Table 7 reports Spearman correlations for all six sub-metric variants against the three ALT primary metrics; WPE with exclusive wins consistently achieves the strongest alignment ($\rho_S = 0.970$–0.996), and the combined $\overline{\mathrm{RP}}$ is reported for the three most informative pairings.

Table 2 summarises all named $\overline{\mathrm{RP}}$ variants.

Table 2. Named $\overline{\mathrm{RP}}$ Variants. All combine RS and WPE as (RS + WPE)/2. Subscripts: excl = exclusive wins, reach = all terminal reaches. Spearman ρ_S against CALT / EALT / AALT from Table 7 ($N = 30$).

Variant	RS source	WPE source	ρ_S (CALT / EALT / AALT)
$\overline{\mathrm{RP}}_{\mathrm{excl}}$	excl	excl	0.970 / 0.996 / 0.951
$\overline{\mathrm{RP}}_{\mathrm{reach}}$	reach	reach	0.967 / 0.926 / 0.982
$\overline{\mathrm{RP}}_{\mathrm{RS\text{-}mxAE}}$	reach	excl	0.948 / 0.902 / 0.959
$\overline{\mathrm{RP}}_{\mathrm{RS\text{-}mxAX}}$	excl	reach	0.952 / 0.981 / 0.924
FRP (weighted, $w_i = 1/n$)	excl	excl	equal to $\overline{\mathrm{RP}}_{\mathrm{excl}}$
ERP (weighted, w_i free)	excl	excl	heterogeneous targets

Fair and Equitable Variants. Fair RP (FRP) uses uniform ideal values $r_i^* = n - 1$ and $t_i^* = v/n$ for all agents, appropriate when equal treatment is desired. *Equitable RP* (ERP) uses the weighted variants RS_i^w and a priority-adjusted WPE with $t_i^* = w_i \cdot v$, enabling deliberate asymmetric allocations directly analogous to weighted proportional fair division in the TEAC sense.

Computational complexity. Computing $\overline{\mathrm{RP}}$ requires a single forward pass over the episode sequence to identify each agent's win episodes ($O(v)$), followed by per-agent gap calculations ($O(n)$ additional steps). The total complexity is $O(v + n)$, a strict improvement over ALT's $O(v \cdot n)$.

4.4 Coordination Score

To assess whether a policy achieves coordination *above chance*, we define a normalised coordination score for any metric M:

$$\mathrm{CS}(M) = \frac{M_{\mathrm{QL}} - M_{\mathrm{rand}}}{1 - M_{\mathrm{rand}}},$$

where M_{QL} is the metric value under Q-learning and M_{rand} under a purely random policy [15]. A positive CS indicates above-chance coordination; a negative CS indicates *worse* coordination than a random baseline.

5 Analytical Comparison of ALT and RP

5.1 Asymptotic Complexity

Let ν denote total episodes and n the agent count. ALT requires $(\nu - n + 1)$ sliding-window passes, each examining n entries. Its time complexity is therefore $\Theta(\nu n)$. RP performs a single sweep of the episode array and constant per-agent post-processing, yielding $\Theta(\nu + n)$. For the typical regime $\nu \gg n$ the speedup factor approaches n, since $(\nu n)/(\nu + n) \approx n$. In practice we observe speedups of 12–25× (see Section 6), growing roughly linearly with n as the theory predicts, though additional implementation overhead (Python function calls, data collection for six parallel variants) raises the constant factor above the theoretical minimum.

5.2 Sensitivity and Expressiveness

The two families differ in what temporal patterns they can detect:

Alternation within windows (ALT advantage). ALT evaluates *which* agents win within each n-episode window. It can therefore distinguish, for example, between a scenario where one agent monopolises all wins in early windows and a scenario where wins are spread evenly from the start. RP, working with per-agent statistics aggregated over the entire sequence, is less sensitive to this temporal structure.

Rhythmic consistency (RP advantage). RP directly measures how close each agent's inter-win gap is to the ideal $n - 1$. This captures scenarios where wins are evenly distributed over time but come in irregular bursts; such a pattern may score moderately on CALT but poorly on RS.

RS vs. threshold-based AWE. Prior work [14] used AWE in place of RS, with a hard threshold at $\bar{r}_i = 2(n-1)$: any agent with a mean wait above this value received AWE = 0, collapsing the rhythm signal for all agents with $n \geq 3$ in our experiments and reducing $\overline{\mathrm{RP}}$ to $\overline{\mathrm{WPE}}/2$. RS eliminates this collapse by degrading continuously: for any finite $\bar{r}_i > r_i^*$ the score is $r_i^*/\bar{r}_i > 0$, preserving a non-trivial rhythmic signal regardless of how infrequently agents win. RS also removes the asymmetry of AWE (penalising under-winning more harshly than over-winning), giving equal penalties to symmetric deviations about the ideal gap.

Multi-agent scaling (RP advantage). For $n = 10$ with $\nu = 385{,}000$ episodes, ALT computation takes approximately 17 seconds while RP completes in under 0.7 seconds. This gap continues to widen with n, making RP the only practical choice for agent populations beyond ~ 50.

5.3 Complementarity

The two families are best understood as complementary rather than competing. CALT provides rich, window-level discrimination and is the primary choice for detailed coordination analysis in small systems. RP provides an efficient, always-computable signal suitable for large-scale simulations,

real-time monitoring, and preliminary screening. In systems where both are tractable, using them jointly provides stronger diagnostic coverage than either alone.

We formalise this as a recommendation:

- For $n \leq 8$ and moderate v: use the three primary ALT metrics (CALT, EALT, AALT) for detailed coordination analysis, with $\overline{\text{RP}}$ as a lightweight cross-check.
- For $n > 8$ or real-time applications: use $\overline{\text{RP}}$ as the primary metric, supplemented by traditional E and RF.
- For fairness auditing distinguishing dominance from exclusivity (e.g., CALT vs. EALT vs. AALT): use the three primary ALT metrics; the auxiliary variants (FALT, qFALT, qEALT) provide additional discrimination near-perfect alternation.

6 Experimental Study

6.1 Setup

We run all experiments under the episode scaling formula derived from the state-space complexity of MBoE [15]:

$$v = B \cdot \left(\frac{n}{2}\right)^2 \cdot \left(1 + \ln \frac{n!}{2!}\right), \quad B = 1000.$$

This yields the episode counts shown in Table 3. For each agent count and reward type (ILF and IQF) we run the Q-learning experiment for both Type-A and Type-B states. Random baselines use a fixed budget of 10,000 episodes each. All experiments were run on an Intel Xeon E5-2640 v4 server (20 cores, 32 GB RAM) running Ubuntu 22.04; computation-time measurements are single-threaded.

Table 3. Episode Budgets by Agent Count

n	Episodes v	RL Runtime	Status
2	4,000	~5 min	Complete
3	9,441	~15 min	Complete
5	31,839	~1 hour	Complete
8	174,583	~6 hours	Complete
10	385,000	~20 hours	Complete

6.2 Q-Learning vs. Random Policy: Detecting Coordination Failure

Table 4 reports the core metrics for Type-A states with ILF rewards, averaged across both random seeds. The pattern is consistent across all configurations (Type-B, IQF) and we report the full data in an online supplement.

Table 4. Metric Values for Q-Learning and Random Policies (Type-A, ILF)

	Q-Learning				**Random**			
n	E	RF	$\overline{\text{RP}}$	CALT	E	RF	$\overline{\text{RP}}$	CALT
2	0.666	0.490	0.538	0.315	0.818	0.972	0.687	0.486
3	0.517	0.921	0.114	0.134	0.866	0.972	0.488	0.359
5	0.457	0.963	0.047	0.059	0.727	0.954	0.242	0.243
8	0.402	0.993	0.015	0.025	0.526	0.893	0.138	0.147
10	0.409	0.989	0.007	0.016	0.443	0.913	0.098	0.111

Several observations stand out. First, Reward Fairness is *high* (> 0.9) for Q-learning agents with $n \geq 3$, which could naïvely be interpreted as successful fair coordination. However, RP and all three primary ALT metrics (CALT, EALT, AALT) tell a different story: these values are substantially *lower* for Q-learning than for the random baseline, indicating that the agents have not learned to coordinate meaningfully.

Second, coordination scores are negative across the board, as shown in Table 5. The Q-learning agents systematically fail to do better than chance at temporal fair division.

Table 5. Coordination Scores: Q-Learning vs. Random Baseline (Type-A, ILF). $\mathrm{CS}(M) = (M_{\mathrm{QL}} - M_{\mathrm{rand}})/(1 - M_{\mathrm{rand}})$. All entries are negative: Q-learning coordinates *worse than chance*.

n	CS(RP)	CS(CALT)	CS(EALT)	CS(AALT)
2	−47.6%	−33.2%	−37.1%	−24.5%
3	−73.0%	−34.9%	−75.7%	−28.8%
5	−25.7%	−24.3%	−54.7%	−20.8%
8	−14.3%	−14.3%	−29.7%	−10.4%
10	−10.1%	−10.6%	−20.7%	−6.6%

The magnitude of the coordination gap decreases as n grows, but for a counter-intuitive reason: the random *baseline* itself improves. With many agents, purely random resource allocation approximates uniform share distribution by the law of large numbers, so the random policy naturally approaches the PA ideal at large n — raising the bar that Q-learning must clear. The gap peaks at $n = 3$ (−73% on RP), where random agents already alternate reasonably well by chance yet Q-learning still fails to coordinate, and shrinks to −10% at $n = 10$ as both policies approach similar low-frequency regimes. At $n = 2$, Q-learning converges to a monopoly (one agent always wins), so $\overline{\mathrm{RP}} = 0.538$ is lower than random (0.687) and RF drops to 0.49; here traditional metrics also detect the failure, making $n = 2$ a distinct case.

6.3 RS Replaces AWE: Eliminating Collapse and Asymmetry

In the conference precursor [14], the rhythm sub-measure was AWE with a hard threshold at $\bar{r}_i = 2(n-1)$. In the present experiments, $\overline{\mathrm{AWE}} = 0$ for *all* configurations with $n \geq 3$ (both Q-learning and random), because agents win infrequently and their mean gaps $\bar{r}_i$ exceed $2(n-1)$ by large margins. Under AWE, this reduces $\overline{\mathrm{RP}}$ to $\overline{\mathrm{WPE}}/2$, losing the rhythmic dimension entirely.

RS eliminates this collapse. Even when $\bar{r}_i \gg r_i^*$, the RS value is $r_i^*/\bar{r}_i$, which remains positive and decreasing: a larger gap yields a lower (not zero) score, and the derivative is everywhere non-zero. The only case where RS = 0 is $k_i < 2$ (agent wins at most once), which correctly represents total coordination failure. This continuous behaviour also means that for configurations where $n = 10$ and $\bar{r}_i \approx 30 \cdot (n-1)$, the RS value is $\approx 1/30$ rather than 0, preserving the rank ordering across configurations that AWE would flatten.

Crucially, RS is also *symmetric*. The former AWE formula $1 - |\bar{r}_i - r_i^*|/r_i^*$ is linear in the deviation and clips over-winning (small $\bar{r}_i$) at the same rate as under-winning (large $\bar{r}_i$) only within the feasible range; beyond $2r_i^*$, under-winning is silently set to 0 regardless of degree. RS assigns equal scores to $\bar{r}_i = c \cdot r_i^*$ and $\bar{r}_i = r_i^*/c$ for any $c > 1$, satisfying the natural requirement that deviating twice as fast in either direction is equally bad.

6.4 Scalability: Computation Time Analysis

Table 6 reports wall-clock computation times for RP versus the full ALT family across all agent counts with their corresponding episode budgets.

Table 6. Computation Time: RP vs. ALT (all six variants), Single Thread

n	Episodes ν	RP (s)	ALT (s)	Speedup	RP% of Total
2	4,000	0.0019	0.023	12×	7.5%
3	9,441	0.0039	0.067	17×	5.3%
5	31,839	0.031	0.672	22×	4.2%
8	174,583	0.245	6.09	25×	3.8%
10	385,000	0.672	17.01	25×	3.7%

The speedup grows with n, from 12× at $n = 2$ to approximately 25× at $n \geq 8$, consistent with the $O(n)$ asymptotic prediction for the ratio $\nu n/(\nu + n) \approx n$. The observed values exceed the bare theoretical minimum of n because ALT accumulates additional overhead from six co-computed variants and per-batch Python function calls that are not captured by asymptotic analysis alone. Importantly, between $n = 8$ and $n = 10$ the speedup appears to plateau; this likely reflects a regime in which both algorithms become dominated by memory-bandwidth costs for the large episode sequences ($\nu \geq 174\,583$), making the asymptotic comparison less informative at this scale.

Figure 1 visualises the computation time scaling on a logarithmic axis; the bar chart on the right panel makes the speedup visually clear.

The practical implication is clear: for $n = 10$ with $\nu = 385{,}000$ episodes, RP completes in under one second whereas ALT requires 17 seconds. For larger n where the episode count continues to grow faster than linearly (owing to the $n!$ factor in the scaling formula), the gap will continue to widen, making RP the only tractable choice for agent populations beyond ≈ 20.

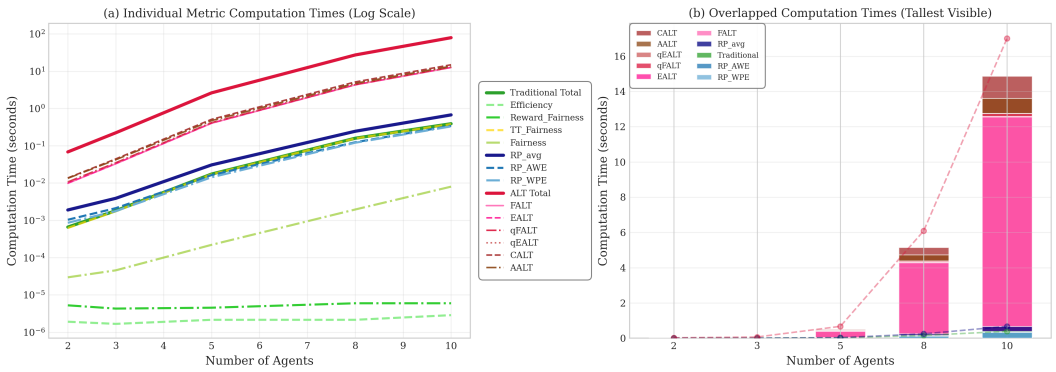

Fig. 1. Wall-clock computation time for RP and the full ALT family as a function of agent count $n \in \{2, 3, 5, 8, 10\}$. Left: individual metric times on a log scale, showing RP (blue) consistently 12–25× faster than any ALT variant. Right: total times by category, making the speedup visually clear. Episode budgets follow Table 3.

6.5 Metric Correlation Analysis

We examine Spearman rank correlations between all RP sub-metric variants (exclusive-win and all-reaches versions of AWE, WPE, and RS, plus three combined $\overline{\text{RP}}$ pairings) and the three ALT primary metrics (CALT, EALT, AALT) across $N = 30$ configurations (20 Q-learning runs: $n \in \{2, 3, 5, 8, 10\}$, 2 state types, 2 reward types; plus 10 random baselines, one per (n, reward) pair). Asymptotic

standard errors ASE = $\sqrt{(1-\rho_S^2)/(N-2)}$ do not exceed 0.063 for any entry with $\rho_S \geq 0.937$ and are omitted from the table for compactness.

Table 7. Spearman ρ_S between RP sub-metric variants and ALT primary metrics ($N = 30$ configurations, all Q-learning and random runs combined). Superscripts: * $p < 0.001$; † $p < 0.05$; $^!$ not significant. AWE_{excl} is constant (= 0) for all Q-learning agents; the moderate ρ_S reflects variation in random baselines only.

Metric	CALT	EALT	AALT
Sub-metrics — exclusive wins			
AWE_{excl}	0.591^*	0.591^*	0.591^*
WPE_{excl}	$\mathbf{0.970}^*$	$\mathbf{0.996}^*$	$\mathbf{0.951}^*$
RS_{excl}	0.937^*	0.970^*	0.905^*
Sub-metrics — all terminal reaches			
$\text{AWE}_{\text{reach}}$	$0.356^!$	$0.307^!$	$0.456^\dagger$
$\text{WPE}_{\text{reach}}$	0.941^*	0.937^*	0.913^*
RS_{reach}	0.718^*	0.669^*	0.789^*
Combined $\overline{\text{RP}} = (\text{RS} + \text{WPE})/2$			
$\text{RS}_{\text{excl}} + \text{WPE}_{\text{excl}}$	0.970^*	0.996^*	0.951^*
$\text{RS}_{\text{reach}} + \text{WPE}_{\text{excl}}$	0.948^*	0.902^*	0.959^*
$\text{RS}_{\text{reach}} + \text{WPE}_{\text{reach}}$	0.967^*	0.926^*	0.982^*

Several patterns stand out.

WPE with exclusive wins dominates all sub-metrics. WPE_{excl} achieves $\rho_S = 0.970$, 0.996, and 0.951 against CALT, EALT, and AALT respectively — the strongest alignment among all standalone sub-metrics. This is expected: both WPE and the EALT/CALT metrics track whether each agent achieves the right *count* of successful episodes relative to the PA ideal, via different computational paths (per-agent frequency count vs. sliding-window batch scoring).

AWE with exclusive wins collapses for Q-learning agents. $\text{AWE}_{\text{excl}} = 0$ for every Q-learning configuration: exclusive wins are too sparse for $n \geq 2$ QL agents (inter-win gaps exceed $2r_i^*$), triggering the hard zero in the original AWE formula. The moderate $\rho_S = 0.591$ in Table 7 is driven entirely by variation in the random baselines. RS replaced AWE in this work precisely to avoid this collapse: RS_{excl} achieves $\rho_S = 0.937$ (CALT) and 0.970 (EALT) by treating gap deviations symmetrically with no hard cutoff.

All-reaches variants show the opposite collapse. $\text{AWE}_{\text{reach}}$ is not significantly correlated with CALT or EALT ($p > 0.05$): the rhythm signal is obscured once simultaneous arrivals inflate the reach count. $\text{WPE}_{\text{reach}}$ is more robust ($\rho_S \approx 0.94$) but still weaker than WPE_{excl}. RS_{reach} ($\rho_S = 0.718$–0.789) is the weakest of the three rhythm sub-metrics, as reaching-with-ties dilutes the per-agent gap signal.

Combined $\overline{\text{RP}}$ *is dominated by the frequency component.* Both the default pairing ($\text{RS}_{\text{excl}} + \text{WPE}_{\text{excl}}$) and the mixed variant ($\text{RS}_{\text{reach}} + \text{WPE}_{\text{excl}}$) achieve $\rho_S \geq 0.948$ against all three ALT metrics. Within Q-learning agents only ($N = 20$), the mixed variant $(\text{RS}_{\text{reach}} + \text{WPE}_{\text{excl}})/2$ reaches $\rho_S = 0.994$ against CALT: reach-based RS provides a non-trivial rhythm signal where AWE would collapse, while WPE on exclusive wins retains the sole-victory-frequency signal. The $\text{RS}_{\text{reach}} + \text{WPE}_{\text{reach}}$ pairing is best for AALT in the full $N = 30$ sample ($\rho_S = 0.982$).

Although the sample spans only five distinct agent counts, the monotone relationship is consistent across all sub-groups (Q-learning vs. random, Type-A vs. Type-B, ILF vs. IQF). The correlations between $\overline{\text{RP}}$ and traditional metrics are substantially weaker ($r \approx 0.31$ with Efficiency, $r \approx 0.18$

with Reward Fairness), confirming that RP captures a distinct temporal dimension not reflected in standard metrics.

Figure 2 illustrates the divergence between temporal fairness metrics (CALT, RP) and the traditional Reward Fairness metric across all agent counts.

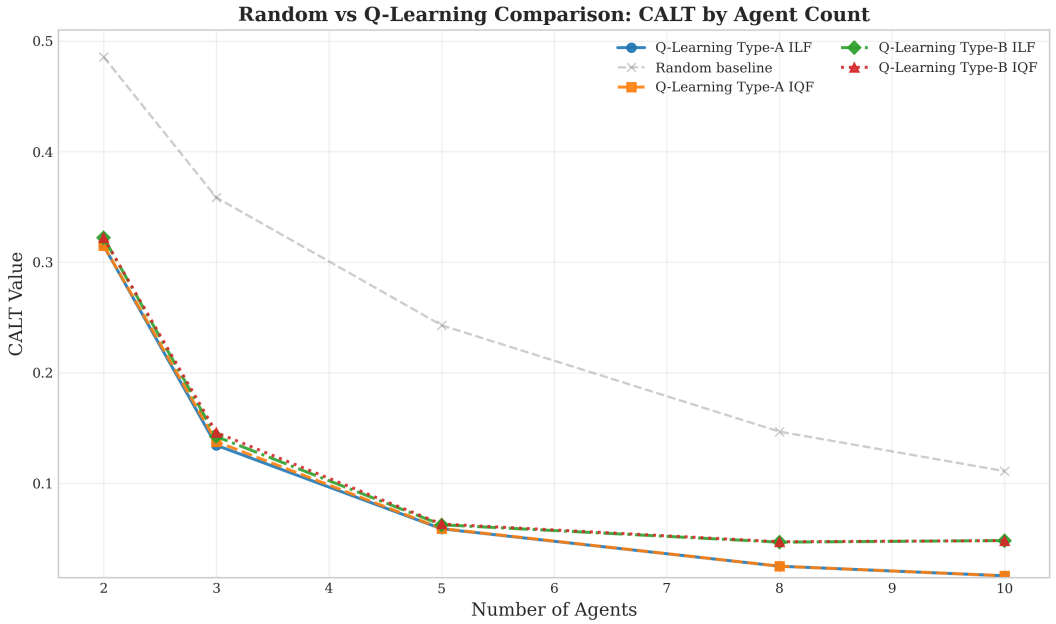

Fig. 2. CALT values for Q-learning (all four configurations: Type-A/B × ILF/IQF) and the random baseline across agent counts $n \in \{2, 3, 5, 8, 10\}$. Q-learning agents (solid coloured lines) fall consistently *below* the random baseline (grey dashed), revealing coordination failure undetectable by traditional metrics.

RS as a scalable proxy for CALT.. RS_{excl} can be viewed as a lightweight approximation of CALT's per-agent coordination signal: rather than computing $O(vn)$ sliding-window batch scores, RS estimates each agent's proximity to the PA gap target in $O(k_i)$ time and combines results into $\overline{\text{RP}}$ in $O(v + n)$ overall. Despite this simplicity, RS_{excl} tracks CALT at $\rho_S = 0.937$ and EALT at $\rho_S = 0.970$; the combined $\overline{\text{RP}}$ (RS_{excl} + WPE_{excl}) achieves $\rho_S = 0.970$ (CALT), 0.996 (EALT), and 0.951 (AALT). In other words, the fine-grained window scoring of CALT does not reveal substantially different coordination failures from what the $O(v + n)$ RP metric already detects, at least within the $n \leq 10$ regime studied here.

6.6 Summary of Empirical Findings

(1) **Coordination failure is widespread.** Q-learning agents fail to achieve temporal fair division in all tested configurations, performing below random baselines on both RP and CALT by 7–73% (peaking at $n = 3$).
(2) **Traditional metrics are misleading.** Reward Fairness exceeds 0.9 even when coordination scores are strongly negative, confirming that temporal fairness metrics are indispensable for accurate evaluation.

(3) **RP sub-metrics align with ALT.** $\overline{\text{RP}}$ (RS_{excl} + WPE_{excl}) achieves $\rho_S = 0.970$, 0.996, and 0.951 against CALT, EALT, and AALT respectively ($N = 30$); WPE with exclusive wins alone reaches $\rho_S = 0.996$ with EALT, validating RP as a scalable proxy for the entire ALT family.
(4) **RP scales gracefully.** The computation time advantage of RP grows with n and v, reaching a 25× speedup at $n = 10$, making it the only tractable option for large systems.
(5) **RS provides a continuous, symmetric rhythm signal.** Unlike AWE, RS does not collapse to zero for $n \geq 3$: it degrades gracefully as $\bar{r}_i / r_i^*$ grows, preserving rank-order discrimination across configurations that the former hard-threshold formula would flatten to zero.

7 Discussion

7.1 Implications for Temporal Fair Division Theory

Our results suggest that learning agents not only fail to converge to Perfect Alternation (the temporally fair solution) but actually produce *worse* outcomes than random resource competition. This is a counterintuitive finding: one might expect that learning would at least match random performance as a lower bound. The explanation lies in the asymmetry of exploration: during early training, ε-greedy Q-learning cycles through policies that lead to occasional wins but does not discover the correlation structure needed for turn-taking. As ε decays, agents lock into strategies that are locally optimal but globally suboptimal, a manifestation of the well-known coordination failure in multi-agent Q-learning [19].

From a fair division perspective, this result has a normative implication: the PA regime (the temporally proportional and envy-free solution) is not self-enforcing under independent learning. Mechanism design interventions (such as reward shaping based on $\overline{\text{RP}}$ or the primary ALT metrics) or explicit coordination protocols may be required to guide agents towards temporal fairness.

7.2 Practical Guidance on Metric Selection

The framework presented in this paper supports a decision process for metric selection depending on available resources and the type of analysis required:

Screening. Use $\overline{\text{RP}}$ (and its sub-components RS, WPE) to quickly flag coordination failures in large or long-running simulations.

Diagnosis. When a failure is detected, apply the three primary ALT metrics (CALT, EALT, AALT) to identify whether the failure is due to temporal monopoly, insufficient exclusivity (EALT), irregular access patterns (CALT), or low per-window coverage (AALT). AltRatio $\approx \sqrt{\text{CALT}}$ maps approximately to the PA-equivalent fraction (the intercept is near-zero [15]); for EALT and AALT the mapping is direct (no square root).

Policy evaluation. For research contexts comparing multiple agent strategies, report CALT, EALT, and AALT as the primary coordination metrics alongside the Coordination Score against a random baseline.

Large-scale deployment. For systems with $n > 10$ or real-time constraints, use $\overline{\text{RP}}$ exclusively; supplement with RF and E to confirm that high RP scores reflect genuine coordination rather than structural artefacts.

7.3 Connection to Mechanism Design

The RP and ALT metrics can serve not only as evaluation tools but also as optimisation objectives. Incorporating $\overline{\text{RP}}$ into reward shaping encourages agents to maintain waiting gaps close to $n-1$ and to access the resource at the ideal frequency v/n. This directly incentivises temporal proportionality. Similarly, maximising CALT as a shaped reward should promote exclusive wins and discourage simultaneous arrivals. We leave the systematic study of RP-based reward shaping as future work,

noting that this direction connects naturally to the mechanism design literature on incentivising fair behaviour [12].

7.4 Limitations

The current framework has two principal limitations. First, the Type-B (memory-augmented) state representation contains an unresolved bug in winner-flag propagation that may affect Type-B results; all primary conclusions are based on Type-A. Second, the episode budget formula assumes identical complexity scaling across all configurations; for very large n this may underestimate the effective state-space size. The AWE collapse identified in the conference precursor [14] is resolved by the RS sub-measure introduced here.

8 Conclusion

We have presented a comprehensive framework for measuring temporal fairness in repeated multi-agent resource competition, grounding it in fair division theory through the connection between Perfect Alternation and temporal proportionality/envy-freeness. The framework spans four complementary layers: traditional metrics (Efficiency, Reward Fairness) that capture aggregate outcomes; the ALT family of sliding-window metrics that capture detailed coordination quality; and Rotational Periodicity, a linear-time metric decomposing temporal fairness into rhythmic and distributional dimensions.

Empirically, we showed that Q-learning agents consistently fail to achieve temporal fair division, performing worse than random baselines by wide margins on RP and all three primary ALT metrics (CALT, EALT, AALT), while maintaining high Reward Fairness—a finding invisible to traditional metrics alone. We also characterised the computational trade-off between the two metric families, confirming that RP achieves 12–25× speedups over ALT with near-perfect rank correlation ($\rho_S \geq 0.95$) at the system level.

We hope these results motivate the inclusion of temporal fairness metrics in evaluation pipelines for multi-agent resource allocation — alongside the classical static fair division measures that currently dominate the literature — since standard metrics, as demonstrated here, actively conceal coordination failure.

Future directions include: (i) developing reward-shaping mechanisms based on RP to guide agents towards temporal fair division; (ii) validating the RS–CALT correlation for larger agent populations ($n > 10$, reduced episode budgets) where ALT becomes intractable; (iii) extending the framework to heterogeneous agents with non-uniform access priorities via the weighted RS and ERP variants; and (iv) applying the framework to real-world scheduling and resource allocation benchmarks.

Acknowledgments

The authors acknowledge the use of Anthropic's Claude AI assistant for editorial assistance and partial mathematical notation formatting. All scientific content, experimental design, results, and conclusions are the exclusive intellectual contribution of the authors, who bear full responsibility for the manuscript. The authors declare no competing interests relevant to this work.

References

[1] Bouveret, S. and Lang, J. (2011). A general elicitation-free protocol for allocating indivisible goods. In *Proc. 22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 73–78. AAAI Press.
[2] Brams, S.J. and Taylor, A.D. (1996). *Fair Division: From Cake-Cutting to Dispute Resolution*. Cambridge University Press.
[3] Caragiannis, I., Kurokawa, D., Moulin, H., Procaccia, A.D., Shah, N., and Wang, J. (2019). The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation*, 7(3):12:1–12:32.

[4] De Jong, S., Tuyls, K., and Verbeeck, K. (2008). Artificial agents learning human fairness. In *Proc. 7th AAMAS*, volume 2, pages 845–852.

[5] Foley, D. (1967). Resource allocation and the public sector. *Yale Economic Essays*, 7:45–98.

[6] Hawkins, R.X.D. and Goldstone, R.L. (2016). The formation of social conventions in real-time environments. *PLOS ONE*, 11(3):e0151670.

[7] He, J., Procaccia, A.D., Psomas, A., and Zeng, D. (2019). Achieving a fairer future by changing the past. In *Proc. 28th IJCAI*, pages 343–349.

[8] Izmirlioglu, Y., Pham, L., Son, T.C., and Pontelli, E. (2024). A survey of multi-agent systems for smart grids. *Energies*, 17(15):3620.

[9] Leibo, J.Z., Zambaldi, V., Lanctot, M., Marecki, J., and Graepel, T. (2017). Multi-agent reinforcement learning in sequential social dilemmas. In *Proc. 16th AAMAS*, pages 464–473.

[10] Lipton, R.J., Markakis, E., Mossel, E., and Saberi, A. (2004). On approximately fair allocations of indivisible goods. In *Proc. 5th ACM EC*, pages 125–131.

[11] Mota, M.P. (2024). *Protocol Emergence with Multi-Agent Reinforcement Learning*. PhD thesis, Université de Lyon.

[12] Moulin, H. (2003). *Fair Division and Collective Welfare*. MIT Press.

[13] Papadopoulos, N.Al. and Sanchez-Fibla, M. (2021). Alternation measures for the evaluation of selfish agents' turn-taking. In *Artificial Intelligence Research and Development*, IOS Press, pages 278–281. DOI: 10.3233/FAIA210145.

[14] Papadopoulos, N.Al., Taratori, R., Sánchez-Fibla, M., and Psannis, K.E. (2025). Rotational Periodicity: A Scalable Metric for Turn-Taking Evaluation in Multi-Agent Systems. In *Proc. 22nd Int. Conf. on Modelling Decisions for Artificial Intelligence (MDAI 2025)*, Valencia, Spain (ISBN 978-91-531-0240-3). Published in Springer LNCS Vol. 15950, DOI: 10.1007/978-3-032-03711-4_16, via publisher administrative error in ICAISC 2025 proceedings.

[15] Papadopoulos, N.Al. and Psannis, K.E. (2026). The coordination gap: Multi-agent alternation metrics for temporal fairness in repeated games. arXiv preprint arXiv:2603.05789. Submitted to *Mathematical Social Sciences*.

[16] Perolat, J., Leibo, J.Z., Zambaldi, V., Beattie, C., Tuyls, K., and Graepel, T. (2017). A multi-agent reinforcement learning model of common-pool resource appropriation. In *Advances in Neural Information Processing Systems 30*, pages 3644–3653.

[17] Raffensperger, P.A., Webb, R.Y., Bones, P.J., and McInnes, A.I. (2011). A simple metric for turn-taking in emergent communication. University of Canterbury Technical Report.

[18] Rankin, D.J., Bargum, K., and Kokko, H. (2007). The tragedy of the commons in evolutionary biology. *Trends in Ecology & Evolution*, 22(12):643–651.

[19] Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge University Press.

[20] Steinhaus, H. (1948). The problem of fair division. *Econometrica*, 16(1):101–104.