IMPERIAL

IMPERIAL COLLEGE LONDON

DEPARTMENT OF MATHEMATICS

DYNAMICS OF LEARNING AND ITERATED GAMES PROJECT 3: REGRET MINIMISATION LEARNING, A COMPUTER IMPLEMENTATION

Learning and Equilibrium Structure in Discrete Colonel Blotto: Regret-Matching Implementation with Symmetrisation

---

CID: 02017478

January 6, 2026

Abstract

This report studies a discrete Colonel Blotto game and connects two complementary perspectives in Discrete Colonel Blotto and General Lotto Games (Hart) [1] and An Introduction to Counterfactual Regret Minimization (Neller–Lanctot)[2]:

(i) Hart's analytic characterisation of equilibrium consistent marginal in symmetric discrete Blotto games via Colonel Lotto, General Lotto, and feasibility constructions;

and (ii) regret-based learning as a computational method to approximate equilibrium play.

We focus on the finite instance with $A = B = 5$ soldiers and $K = 3$ battlefields, which yields a 21×21 zero-sum matrix game. We implement regret-matching in self-play and verify convergence in a set-robust way using exploitability, value convergence, and the induced marginal troop distribution, which we compare against Hart's analytical description of the equilibrium set in the symmetric discrete setting.

Contents

1 Introduction – The Colonel Blotto Game

Overview. The **Colonel Blotto game** is a classic strategic allocation game that appears in political science, economics, military strategy, and more recently in **learning in games**. In this project, we follow the algorithm described in Neller–Lanctot [2], and we connect it to Hart's analytic approach [1] to symmetric discrete Blotto via reductions to Colonel Lotto / General Lotto.

1.1 Discrete Colonel Blotto

Definition 1.1. Fix integers $A, B \geq 1$ (soldiers) and $K \geq 2$ (battlefields/ partitions). Each player chooses a *pure strategy*

$$x = (x_1, \ldots, x_K) \in \mathbb{Z}_{\geq 0}^K \qquad \text{such that} \qquad \sum_{i=1}^{K} x_i = A,$$

$$y = (y_1, \ldots, y_K) \in \mathbb{Z}_{\geq 0}^K \qquad \text{such that} \qquad \sum_{i=1}^{K} y_i = B,$$

interpreted as allocating x_i soldiers to battlefield i. The set of feasible allocations is

$$\mathcal{X}_{A,K} = \left\{ x \in \mathbb{Z}_{\geq 0}^K : \sum_{i=1}^{K} x_i = A \right\}.$$

$$\mathcal{X}_{B,K} = \left\{ y \in \mathbb{Z}_{\geq 0}^K : \sum_{i=1}^{K} y_i = B \right\}.$$

Given allocations $x \in \mathcal{X}_{A,K}$ (Player 1) and $y \in \mathcal{X}_{B,K}$ (Player 2), the outcome is decided battlefield-by-battlefield. Following Hart's convention:

$$\operatorname{sgn}(x_i - y_i) = \begin{cases} 1, & x_i > y_i, \\ 0, & x_i = y_i, \\ -1, & x_i < y_i. \end{cases}$$

The (normalised) payoff to Player 1 is then

$$u(x, y) = \frac{1}{K} \sum_{i=1}^{K} \operatorname{sgn}(x_i - y_i),$$

Interpretation. Player 1 is rewarded for winning battlefields and penalised for losing them, with ties neutral. The factor $\frac{1}{K}$ is a normalisation by the number of battlefields, so that payoffs are scaled to lie in $[-1, 1]$.

Colonel Blotto is a zero sum game. Player 2's payoff is

$$u_2(x, y) = -u(x, y),$$

so any gain for one player is an equal loss for the other.

Example. Let $(A = B, K) = (5, 3)$. A valid allocation is

$$x = (2, 2, 1) \in \mathcal{X}_{5,3},$$

If the opponent plays $y = (3, 1, 1)$, then

$$\big(\operatorname{sgn}(2-3), \operatorname{sgn}(2-1), \operatorname{sgn}(1-1)\big) = (-1, 1, 0), \qquad u(x, y) = 0,$$

therefore both player's payoff is 0 in which the round is a draw.

Strategic allocation. Should the player:

- spread forces evenly to avoid catastrophic losses?
- gamble by heavily investing in a few key fronts?
- bluff by sacrificing one battlefield to dominate others?

These strategic trade-offs make the Colonel Blotto game a rich environment for studying adaptive behaviour. Our aim is to understand how learning dynamics guide players toward strategies that achieve near-optimal payoffs over repeated play.

1.2 Minmax value and Nash equilibrium.

Let $M \in \mathbb{R}^{m \times n}$ be the payoff matrix for the discrete blotto game. Here $m := |\mathcal{X}^{(1)}|$ and $n := |\mathcal{X}^{(2)}|$ denote the numbers of pure strategies available to Player 1 and Player 2, respectively.

Using the lecture note notations [4], Player 1 chooses a mixed strategy $x \in \Delta_m$ over rows, Player 2 chooses a mixed strategy $y \in \Delta_n$ over columns. In addition, Player 1's expected payoff is $x^\top My$ (with Player 2 receiving $-x^\top My$).

Theorem 1.1 (Theorem 4.3: minmax Theorem for finite two-player zero-sum games presented in Dynamics of Learning and Iterated Games: Lecture Notes [4])**.** *There exists a value $v \in \mathbb{R}$ such that*

$$\max_{x \in \Delta_m} \min_{y \in \Delta_n} x^\top My = \min_{y \in \Delta_n} \max_{x \in \Delta_m} x^\top My = v.$$

Moreover, a (mixed) Nash equilibrium exists, and every *Nash equilibrium attains the same value v.*

Remark. Different equilibrium mixed strategies may exist, but they are all payoff-equivalent in the sense that they achieve the optimal minmax value.

In this project, we focus on **symmetric** Blotto games, where both players have the same pure strategy set, so $m = n$ and M is an $n \times n$ matrix. In particular, we study the instance with total resources $A = B = 5$ distributed across $K = 3$ battlefields, the game $B(5, 5; 3)$.

1.3 Project plan.

We first dive into Hart's reduction framework and equilibrium characterisation for *symmetric discrete Blotto* (via Colonel Lotto, General Lotto, and feasibility constructions) in section 2–6. We then implement Neller–Lanctot's regret-matching algorithm in section 8 for the resulting 21×21 zero-sum matrix game, and evaluate convergence in a set-robust way.

2 Hart's reduction pipeline for discrete Blotto

Overview. The goal of Hart's paper [1] is to characterise *optimal play* (Nash equilibrium strategies) in discrete symmetric Colonel Blotto games. The key idea is that Blotto is hard to solve directly because strategies are *high-dimensional allocations* with a combinatorial strategy space.

Remark. Each Blotto strategy specifies how resources are allocated across K battlefields, so pure strategies live in a $(K-1)$-dimensional simplex defined by the budget constraint. The number of feasible integer allocations grows combinatorially with S and K, making the normal-form payoff matrix extremely large and rendering direct best-response or equilibrium computation impractical. This explains why Blotto is difficult to solve directly and why Hart adopts a more structural approach.

Hart's main method is to

i reduce discrete Blotto to simpler auxiliary games (Colonel Lotto, General Lotto) where equilibrium structure can be analysed explicitly

ii map the resulting solutions back to Blotto via a feasibility construction.

In particular, Hart derives:

- an analytic description of the *equilibrium value* and the *form of optimal one-dimensional marginals* for symmetric games.
- explicit *feasibility constraints* ensuring that these marginals can be realised by an actual Blotto mixed strategy.

Outside these regimes, and especially in more general *asymmetric* cases, the full analytic solution is not yet available in the same level of generality.

In this project we focus on the **symmetric** discrete setting (both players have the same resources and the same set of feasible allocations), for which Hart characterises the value and a family of equilibrium-consistent one-dimensional marginals.

3 From Colonel Blotto to Colonel Lotto and General Lotto

Overview. Hart's approach replaces the high-dimensional allocation problem in Blotto by simpler *one-dimensional marginal* games that capture the equilibrium mixed strategy.

In this section we define **Colonel Lotto** and **General Lotto** (both discrete and continuous versions), and explain the Interpretation behind each reduction.

3.1 Discrete Colonel Blotto (recall section 1.1)

Fix integers $A \geq 1$ and $K \geq 2$. A pure Blotto strategy is an allocation $x = (x_1, \dots, x_K) \in \mathbb{Z}_{\geq 0}^K$ with $\sum_{i=1}^K x_i = A$. Given allocations x, y, Hart's normalised payoff to Player 1 is

$$u(x, y) \;=\; \frac{1}{K} \sum_{i=1}^{K} \operatorname{sgn}(x_i - y_i), \qquad \operatorname{sgn}(t) \in \{-1, 0, 1\}.$$

The key difficulty is that Blotto strategies are joint distributions over K coordinates subject with feasibility constraint: nonnegative integers summing to A.

3.2 Colonel Lotto L(A,B;K) (Symmetrised version of Blotto Game)

Definition 3.1 (Colonel Lotto)**.** Fix integers $A, B \geq 0$ and $K \geq 1$. Player 1 chooses a vector $x = (x_1, \ldots, x_K) \in \mathbb{Z}^K_{\geq 0}$ with $\sum_{k=1}^K x_k = A$, and Player 2 chooses $y = (y_1, \ldots, y_K) \in \mathbb{Z}^K_{\geq 0}$ with $\sum_{\ell=1}^K y_\ell = B$. The payoff to Player 1 by comparing *all cross-matchings* of battlefields:

$$h_L(x, y) \;:=\; \frac{1}{K^2} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \operatorname{sgn}(x_k - y_\ell), \qquad \operatorname{sgn}(t) \in \{-1, 0, 1\}.$$

Equivalently, if I and J are independent uniform random indices on $\{1, \ldots, K\}$, then

$$h_L(x, y) \;=\; \mathbb{E}[\operatorname{sgn}(x_I - y_J)]\,.$$

Interpretation. Colonel Lotto keeps the same underlying allocation sets, but changes the evaluation rule: instead of comparing only the K matched pairs (x_k, y_k), it compares *every* pair (x_k, y_ℓ) and averages the result. This averaging collapses the strategic problem towards ***one-dimensional marginal behaviour***.

Using the idea from Hart's paper [1], only the empirical distributions of the coordinates of x and y matter for h_L: if we pick a battlefield uniformly at random from Player 1's allocation, we obtain a random variable X taking values in $\mathbb{Z}_{\geq 0}$ with mean $\mathbb{E}[X] \;=\; \frac{A}{K}$, and similarly for Player 2 we obtain Y with $\mathbb{E}[Y] = \frac{B}{K}$, where A and B are the players' **total** troop budgets, while K is the number of battlefields; the ratios A/K and B/K are the **average troops per battlefield**, which are the natural parameters that will appear in General Lotto.

Example. Take $A = 3$, $B = 2$, $K = 2$. If Player 1 plays $x = (3, 0)$ and Player 2 plays $y = (2, 0)$, then

$$h_L(x, y) = \frac{1}{4}\Big(\operatorname{sgn}(3-2) + \operatorname{sgn}(3-0) + \operatorname{sgn}(0-2) + \operatorname{sgn}(0-0)\Big) = \frac{1}{4}(1 + 1 - 1 + 0) = \frac{1}{4}.$$

3.3 Connection between Blotto and Lotto

Recall the payoff of Colonel Blotto and Colonel Lotto:

$$h_B(x, y) \;:=\; \frac{1}{K} \sum_{k=1}^{K} \operatorname{sgn}(x_k - y_k),$$

$$h_L(x, y) \;:=\; \frac{1}{K^2} \sum_{k=1}^{K} \sum_{\ell=1}^{K} \operatorname{sgn}(x_k - y_\ell).$$

The key symmetrisation idea is to remove labels by randomly permuting the coordinates of a Blotto allocation.

Definition 3.2 (Permutation-mixing (symmetrisation) operator)**.** Let S_K be the set of all $K!$ permutations of $\{1, \ldots, K\}$. For a pure allocation $x = (x_1, \ldots, x_K) \in \mathbb{Z}^K_{\geq 0}$, define the mixed strategy $\sigma(x)$ to be the uniform distribution over all coordinate-permutations of x:

$$\sigma(x) \;:=\; \text{Uniform}\big\{(x_{\pi(1)}, \ldots, x_{\pi(K)}) : \pi \in S_K\big\}.$$

Equivalently,

$$\sigma(x) \;=\; \frac{1}{|S_K|} \sum_{\pi \in S_K} \delta_{\pi x}, \qquad \text{where } \pi x := (x_{\pi(1)}, \ldots, x_{\pi(K)}),$$

and δ_z denotes the point mass at z.

For a mixed strategy ξ on allocations (i.e. a probability distribution on $\mathbb{Z}^K_{\geq 0}$), $\xi(x)$ denotes the probability that ξ assigns to the pure allocation x. Define $\sigma(\xi)$ by applying $\sigma(\cdot)$ to each pure allocation in its support:

$$\sigma(\xi) \;:=\; \sum_x \xi(x)\,\sigma(x).$$

Examples.

(a) *Pure strategy.* Let $K = 3$ and $x = (2, 1, 0)$. All $3! = 6$ permutations are distinct:

$$(2, 1, 0),\ (2, 0, 1),\ (1, 2, 0),\ (1, 0, 2),\ (0, 2, 1),\ (0, 1, 2).$$

Hence $\sigma(x)$ plays each with probability $1/6$.

(b) *Pure strategy with repeats.* Let $K = 3$ and $x = (1, 1, 0)$. There are only 3 distinct permutations:

$$(1, 1, 0),\ (1, 0, 1),\ (0, 1, 1),$$

each arising twice among the 6 permutations, so $\sigma(x)$ assigns probability $2/6 = 1/3$ to each.

(c) *Mixed strategy.* Let $K = 3$ and

$$\xi \;=\; 0.7\,\delta_{(2,1,0)} \;+\; 0.3\,\delta_{(1,1,1)}.$$

Then

$$\sigma(\xi) \;=\; 0.7\,\sigma\big((2, 1, 0)\big) \;+\; 0.3\,\sigma\big((1, 1, 1)\big).$$

Since $(1, 1, 1)$ is already invariant under permutations, $\sigma((1, 1, 1)) = \delta_{(1,1,1)}$. Therefore, each of the 6 permutations of $(2, 1, 0)$ has probability $0.7/6$, and $(1, 1, 1)$ has probability 0.3.

Lemma 3.1 (Symmetrised Blotto equals Lotto)**.** *For every pair of pure strategies x, y,*

$$h_B(\sigma(x), y) \;=\; h_L(x, y).$$

More generally, for every mixed strategies ξ of Player 1 and η of Player 2,

$$h_B(\sigma(\xi), \eta) \;=\; h_L(\xi, \eta), \qquad h_B(\xi, \sigma(\eta)) \;=\; h_L(\xi, \eta),$$

and in particular,

$$h_B(\sigma(\xi), \sigma(\eta)) \;=\; h_L(\xi, \eta).$$

Proof. Fix pure allocations $x, y \in \mathbb{Z}^K_{\geq 0}$.

Step 1. By definition of $h_B(x, y)$ and $\sigma(x)$, Player 1 plays the permuted vector $\pi(x)$ where π is uniform on S_K. Thus

$$h_B(\sigma(x), y) = \mathbb{E}_\pi[h_B(\pi(x), y)] = \mathbb{E}_\pi\left[\frac{1}{K}\sum_{k=1}^{K} \operatorname{sgn}\big((\pi(x))_k - y_k\big)\right] = \mathbb{E}_\pi\left[\frac{1}{K}\sum_{k=1}^{K} \operatorname{sgn}\big(x_{\pi(k)} - y_k\big)\right].$$

Step 2 . By linearity of expectation,

$$h_B(\sigma(x), y) = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_\pi\left[\operatorname{sgn}\big(x_{\pi(k)} - y_k\big)\right].$$

Step 3. By uniformity of $\pi(k)$, for each fixed k, the random index $\pi(k)$ is uniform over $\{1, \dots, K\}$, hence

$$\mathbb{E}_\pi \left[\operatorname{sgn}\big(x_{\pi(k)} - y_k\big)\right] = \frac{1}{K} \sum_{\ell=1}^{K} \operatorname{sgn}(x_\ell - y_k).$$

Step 4. Substituting back and swapping the order of summation,

$$h_B(\sigma(x), y) = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{K} \sum_{\ell=1}^{K} \operatorname{sgn}(x_\ell - y_k) = \frac{1}{K^2} \sum_{\ell=1}^{K} \sum_{k=1}^{K} \operatorname{sgn}(x_\ell - y_k) = h_L(x, y),$$

which is exactly the Colonel Lotto payoff.

Mixed strategies and symmetrising Player 2. The payoff is bilinear in mixed strategies, so the same identity extends by linearity: for any mixed strategies ξ, η,

$$h_B(\sigma(\xi), \eta) = h_L(\xi, \eta), \qquad h_B(\xi, \sigma(\eta)) = h_L(\xi, \eta),$$

and in particular $h_B(\sigma(\xi), \sigma(\eta)) = h_L(\xi, \eta)$. □

Implication. Lemma 3.1 implies that uniformly randomising battlefield labels in Blotto yields the Lotto payoff: for any mixed strategies ξ, η,

$$h_B(\sigma(\xi), \eta) = h_L(\xi, \eta) \quad \text{and} \quad h_B(\xi, \sigma(\eta)) = h_L(\xi, \eta).$$

Hence Lotto is the *symmetrised* version of Blotto. Moreover, if one player plays a permutation-invariant strategy (applies σ), then only the other player's symmetrisation matters for payoffs: any battlefield-label asymmetry is payoff-irrelevant, since replacing η by $\sigma(\eta)$ leaves $h_B(\sigma(\xi), \eta)$ unchanged.

3.4 General Lotto

Definition 3.3 (General Lotto)**.** For real numbers $a, b \geq 0$, the **General Lotto** game $\Gamma(a, b)$ is defined as follows.

Player 1 chooses an integer-valued random variable $X \in \mathbb{Z}_{\geq 0}$ with fixed mean $\mathbb{E}[X] = a$, and Player 2 chooses an integer-valued random variable $Y \in \mathbb{Z}_{\geq 0}$ with fixed mean $\mathbb{E}[Y] = b$.

The payoff to Player 1 is

$$H(X, Y) \;:=\; \mathbb{P}(X > Y) - \mathbb{P}(X < Y) \;=\; \mathbb{E}[\operatorname{sgn}(X - Y)]\,.$$

Ties may occur with positive probability (since X, Y are integer-valued), in which case $\operatorname{sgn}(0) = 0$ and ties contribute 0 to the payoff.

Interpretation. General Lotto is a relaxation of Colonel Lotto in which we keep only the **mean constraints** and discard the restriction that X and Y must arise as the *uniformly drawn coordinate* of a valid K-battlefield allocation. Crucially, in $\Gamma(a, b)$ each player chooses the *entire distribution* of X (and of Y) strategically, subject only to $\mathbb{E}[X] = a$ and $\mathbb{E}[Y] = b$. The parameters a and b represent the players' average resources per battlefield, with

$$a = \frac{A}{K}, \qquad b = \frac{B}{K},$$

linking Colonel Lotto to General Lotto. Later, we return to the feasibility question in section 5: which General Lotto marginals can actually be implemented by valid K-battlefield partitions?

Example. Let $a = 1$ and $b = 1$. Suppose Player 1 chooses

$$X = \begin{cases} 0 & \text{with prob. } \frac{1}{2}, \\ 2 & \text{with prob. } \frac{1}{2}, \end{cases} \qquad (\mathbb{E}[X] = 1),$$

and Player 2 chooses $Y \equiv 1$ (always 1). Then

$$H(X, Y) = \mathbb{P}(X > 1) - \mathbb{P}(X < 1) = \tfrac{1}{2} - \tfrac{1}{2} = 0.$$

3.5 General Lotto (continuous analogue)

Definition 3.4 (Continuous General Lotto)**.** Now players may choose nonnegative real-valued random variables$X \in \mathbb{R}_{\geq 0}, Y \in \mathbb{R}_{\geq 0}$, with fixed means $\mathbb{E}[X] = a$ and $\mathbb{E}[Y] = b$. The payoff remains

$$H(X, Y) \;=\; \mathbb{P}(X > Y) - \mathbb{P}(X < Y) \;=\; \mathbb{E}[\operatorname{sgn}(X - Y)]\,,$$

where ties occur with probability zero under typical continuous distributions.

Interpretation. The continuous version may be viewed as an idealised limit of the discrete model in which troop units become divisible. Hart studies it because it is *mathematically easier*. In the reduction pipeline, the continuous solution provides a useful template for the *structure* of optimal marginals, which can then be discretised and checked for feasibility in the discrete setting.

A further advantage is that existence arguments are cleaner in the continuous formulation. The minmax/Nash value follows from standard compactness and continuity conditions, before returning to the discrete case where integrality and feasibility must be enforced when mapping back to Colonel Lotto and ultimately to Blotto.

Example. Let $a = b = 1$. If Player 1 chooses $X \sim \text{Unif}[0, 2]$ and Player 2 chooses $Y \equiv 1$, then

$$\mathbb{P}(X > 1) = \frac{1}{2}, \qquad \mathbb{P}(X < 1) = \frac{1}{2}, \qquad \Rightarrow \qquad H(X, Y) = 0.$$

4 Hart's General Lotto solutions (continuous and discrete)

Overview. Hart [1] first solves the continuous version of General Lotto (real-valued X, Y with fixed means), which provides a clean benchmark and interpretation for the structure of optimal play. He then focuses on the main objective in solving discrete Blotto where he first solve the *discrete* General Lotto game. In what follows we do not reprove these results. We state the main theorem statements and briefly highlight the Interpretation behind their form.

Recall the General Lotto payoff in subsection 3.4

$$H(X, Y) = \mathbb{P}(X > Y) - \mathbb{P}(X < Y) = \mathbb{E}[\operatorname{sgn}(X - Y)],$$

where Player 1 chooses a nonnegative random variable X with $\mathbb{E}[X] = a$ and Player 2 chooses Y with $\mathbb{E}[Y] = b$.

4.1 Continuous General Lotto Solution.

Hart solves the continuous version $\mathcal{G}(a, b)$ (real-valued X, Y with the same mean constraints).

Theorem 4.1 (Taken directly from [1])**.** *Let $a \geq b > 0$. The value of the continuous General Lotto game is*

$$\text{val } \mathcal{G}(a, b) = \frac{a-b}{a} = 1 - \frac{b}{a}.$$

Moreover, the unique optimal strategies are

$$X^\star \sim U(0, 2a), \qquad Y^\star = \begin{cases} 0, & \textit{with probability } 1 - \dfrac{b}{a}, \qquad \textit{(i)} \\ U(0, 2a), & \textit{with probability } \dfrac{b}{a}. \qquad \textit{(ii)} \end{cases}$$

(Equivalently, $Y^\star \sim \left(1 - \frac{b}{a}\right)\delta_0 + \frac{b}{a} U(0, 2a)$.)

Interpretation. Player 1 (the stronger player when $a \geq b$) plays $X^\star \sim U(0, 2a)$. This makes it hard for Player 2 to counter as they cannot pick a single response that beats X often, because X is equally likely to fall below or above that. In particular, when Player 2 also uses a uniform draw on the same interval, the match-up is perfectly symmetric, so wins and losses cancel on average.

Player 2 cannot afford to match Player 1's mean ($b < a$), so the optimal response is to *mix* between (i) conceding and (ii) matching Player 1's strategy on the same support. With probability $1 - \frac{b}{a}$, Player 2 plays 0 (conceding), and with probability $\frac{b}{a}$, Player 2 draws uniformly from $[0, 2a]$ just like Player 1. The mixing weight $\frac{b}{a}$ is chosen so that the mean constraint is satisfied:

$$\mathbb{E}[Y^\star] = \left(1 - \frac{b}{a}\right) \cdot 0 \; + \; \frac{b}{a} \cdot \mathbb{E}[U(0, 2a)] = \frac{b}{a} \cdot a = b.$$

When Player 2 uses the uniform component, the match-up against $X^\star$ is symmetric so the expected payoff is zero. The only disadvantage comes from the "concession" probability. Each time Player 2 plays 0, Player 1 wins with probability 1 (ties have probability 0 in the continuous model), so the value is exactly the fraction of rounds in which Player 2 must concede: $\text{val } \mathcal{G}(a, b) = 1 - \frac{b}{a}$.

4.2 Discrete General Lotto Solution.

Now X, Y are *integer-valued* with the same mean constraints in Hart's discrete game $\Gamma(a, b)$.

Let $a \geq b > 0$ and for integer $m \geq 0$, Hart defines three uniform distributions (normal, odd, even) with mean m:

$$U_m := U(\{0, 1, \dots, 2m\}), \qquad U_m^o := U(\{1, 3, \dots, 2m-1\}), \qquad U_m^e := U(\{0, 2, \dots, 2m\}),$$

and $\text{conv}\{U_m^o, U_m^e\} := \{\lambda U_m^o + (1-\lambda) U_m^e : \lambda \in [0, 1]\}$.

Discrete Solution case split. To cover all solution cases with $a \geq b > 0$, Hart splits the discrete analysis by (i) whether a is an integer, and (ii) when a is non-integer, by the relative sizes of the integer parts $\lfloor a \rfloor$ and $\lfloor b \rfloor$ (Theorems 2–4).

Figure 1 taken directly from [1] summarises the cases, specifically:

- **Rows 1–2** (cases $a = m$): covered by **Theorem 2**.
  - Row 1: $a = m,\ b = m$ (symmetric integer case).
  - Row 2: $a = m,\ b < m$ (integer a vs smaller b).
- **Row 3** (non-integer a with $\lfloor a \rfloor < \lfloor b \rfloor$): covered by **Theorem 3**.
  - Row 3: $a = m + \alpha,\ b = m + \beta$ with $0 < \alpha < \beta < 1$ (so $\lfloor a \rfloor = m < m + 1 = \lfloor b \rfloor$).
- **Rows 4–5** (non-integer a with $\lfloor a \rfloor \geq \lfloor b \rfloor$): covered by **Theorem 4**.
  - Row 4: $a = m + \alpha,\ b = m$ (same integer part).
  - Row 5: $a = m + \alpha,\ b < m$ (strictly smaller integer part for b).

For the purpose of this project, we will only focus on Row 1.

Remark. For full details of the theorems, see Appendix A.1

	Value	Optimal Strategies		
$a = m$, $b = m$	0	**A**	*	$U^m_{\mathrm{O/E}}$
		B	*	$U^m_{\mathrm{O/E}}$
$a = m$, $b < m$	$\frac{a-b}{a} = 1 - \frac{b}{m}$	**A**	*	U^m_{O}
		B		$\left(1 - \frac{b}{m}\right) \mathbf{1}_0 + \left(\frac{b}{m}\right) U^m_{\mathrm{O/E}}$
$a = m + \alpha$, $b = m + \beta$	$\frac{a-b}{\lceil a \rceil} = \frac{\alpha - \beta}{m+1}$	**A**	*	$(1-\alpha)\, U^m_{\mathrm{E}} + \alpha\, U^{m+1}_{\mathrm{O}}$
		B	*	$(1-\beta)\, U^m_{\mathrm{E}} + \beta\, U^{m+1}_{\mathrm{O}}$
$a = m + \alpha$, $b = m$	$\frac{a-b}{\lceil a \rceil} = \frac{\alpha}{m+1}$	**A**		$(1-\alpha)\, U^m_{\mathrm{O/E}} + \alpha\, U^{m+1}_{\mathrm{O}}$
		B	*	U^m_{E}
$a = m + \alpha$, $b < m$	$(1-\alpha)\frac{\lfloor a \rfloor - b}{\lfloor a \rfloor} + \alpha \frac{\lceil a \rceil - b}{\lceil a \rceil}$	**A**		$(1-\alpha)\, U^m_{\mathrm{O}} + \alpha\, U^{m+1}_{\mathrm{O}}$
		B	*	$\left(1 - \frac{b}{m}\right) \mathbf{1}_0 + \left(\frac{b}{m}\right) U^m_{\mathrm{E}}$

Figure 1 Hart's case table for discrete General Lotto $\Gamma(a, b)$: values and optimal strategies across regimes. A star indicates uniqueness of the optimal strategy (up to equivalence).

5 From General Lotto back to Colonel Blotto: value preservation and feasibility

Overview. With the explicit equilibrium structure for *discrete* General Lotto, Hart's next step is to connect these optimal *marginals* back to the original K-battlefield Colonel Blotto game. Conceptually, I will present this in two layers.

1. Symmetrisation shows that passing from Blotto to Lotto preserves the *minmax value* of the game, so solving Lotto gives the correct value benchmark for Blotto.

2. To recover actual Blotto strategies one must address a feasibility question: *which General Lotto marginal distributions can arise from random K-partitions of total resource A?*

5.1 Value preservation under symmetrisation

Let v_B and v_L denote the minmax values of Blotto and Colonel Lotto, respectively:

$$v_B := \max_{\xi} \min_{\eta} h_B(\xi, \eta), \qquad v_L := \max_{\xi} \min_{\eta} h_L(\xi, \eta).$$

The symmetrisation lemma 3.1 implies that for all mixed strategies ξ, η,

$$h_B(\sigma(\xi), \sigma(\eta)) \;=\; h_L(\xi, \eta).$$

Consequently, the two games have the same value:

$$v_B \;=\; v_L.$$

Remark. This is primarily a statement about *payoffs/values*, not a literal equality of equilibrium sets: Blotto and Lotto need not share the same Nash equilibrium *set*, since Lotto averages away battlefield alignment constraints present in Blotto.

In fact in exchangeable-battlefields setting is that the original Blotto equilibrium set *contains* permutation-invariant (symmetrised) equilibria,

$$\mathrm{NE}_{\mathrm{sym}} \subseteq \mathrm{NE}_{\mathrm{Blotto}}.$$

The map $(p, q) \mapsto (\sigma(p), \sigma(q))$ is generally many-to-one, so this does not identify the full equilibrium set. It identifies a payoff-equivalent, permutation-invariant subclass.

Lemma 5.1 (Value preservation)**.** *The minmax values of Blotto and Colonel Lotto coincide: $v_B = v_L$.*

Proof. Let

$$v_B := \max_{\xi} \min_{\eta} h_B(\xi, \eta), \qquad v_L := \max_{\xi} \min_{\eta} h_L(\xi, \eta).$$

Remark. Since the strategy sets are finite and the payoff is bilinear, von Neumann's minmax theorem applies, so $\max\min = \min\max$.

Step 1: Show $v_B \geq v_L$. Fix any Lotto mixed strategy ξ. Consider the Blotto mixed strategy $\sigma(\xi)$ obtained by randomly relabelling the battlefields before play. By the symmetrisation lemma 3.1, for every opponent strategy η,

$$h_B(\sigma(\xi), \eta) = h_L(\xi, \eta).$$

In words: for any play ξ in Lotto, Player 1 can play $\sigma(\xi)$ in Blotto and obtain exactly the same expected payoff against any opponent. Therefore,

$$\min_{\eta} h_B(\sigma(\xi), \eta) = \min_{\eta} h_L(\xi, \eta).$$

Taking $\max_{\xi}$ over Player 1's choices yields

$$v_B = \max_{\xi'} \min_{\eta} h_B(\xi', \eta) \;\geq\; \max_{\xi} \min_{\eta} h_B(\sigma(\xi), \eta) = \max_{\xi} \min_{\eta} h_L(\xi, \eta) = v_L.$$

In words: any maximin payoff that Player 1 can guarantee in Lotto can also be guaranteed in Blotto (by symmetrising), hence $v_B \geq v_L$.

Step 2: Show $v_B \leq v_L$. Fix any Blotto mixed strategies ξ, η. Symmetrising Player 2's strategy does not change the payoff:

$$h_B(\xi, \sigma(\eta)) = h_L(\xi, \eta) \qquad \text{(by the same lemma 3.1, symmetrising Player 2)}.$$

Hence, for each fixed ξ,

$$\min_\eta h_B(\xi, \eta) \;\leq\; \min_\eta h_B(\xi, \sigma(\eta)) = \min_\eta h_L(\xi, \eta).$$

Taking $\max_\xi$ over Player 1's choices gives

$$v_B = \max_\xi \min_\eta h_B(\xi, \eta) \;\leq\; \max_\xi \min_\eta h_L(\xi, \eta) = v_L.$$

Combining Step 1 and Step 2 yields $v_B = v_L$. □

5.2 Feasibility

Fix integers $A \geq 1$ and $K \geq 2$. For a K-partition $x = (x_1, \ldots, x_K) \in \mathbb{Z}^K_{\geq 0}$ of A ($\sum_{k=1}^K x_k = A$), associate the induced one-coordinate discrete random variable

$$X_x \;:=\; \sum_{k=1}^{K} \frac{1}{K}\, \mathbf{1}_{\{x_k\}} \qquad , X_x = x_I \text{ where } I \sim \mathrm{Unif}\{1, \ldots, K\}$$

Then $\mathbb{E}[X_x] = A/K$.

Definition 5.1 ((A, K)-feasible (Hart))**.** A nonnegative integer-valued random variable Z is called (A, K)-*feasible* if it can be obtained from a probability distribution over K-partitions of A.

Equivalently, Z is (A, K)-feasible if there exist K-partitions $x^{(1)}, \ldots, x^{(n)}$ of A and weights $\lambda_1, \ldots, \lambda_n > 0$ with $\sum_{i=1}^n \lambda_i = 1$ such that

$$Z \;\stackrel{d}{=}\; \sum_{i=1}^{n} \lambda_i \, X_{x^{(i)}}.$$

(In words: Z is a convex combination of the coordinate-marginals generated by valid K-partitions of A.)

Example. Let $(A, K) = (10, 4)$. Consider the distribution Z on $\{0, 1, 2, 3, 4, 10\}$ given by

$$Z \sim \frac{3}{8}\mathbf{1}_{\{0\}} + \frac{1}{8}\mathbf{1}_{\{10\}} + \frac{1}{8}\mathbf{1}_{\{1\}} + \frac{1}{8}\mathbf{1}_{\{2\}} + \frac{1}{8}\mathbf{1}_{\{3\}} + \frac{1}{8}\mathbf{1}_{\{4\}}.$$

We claim that Z is $(10, 4)$-feasible.

Indeed, take the two pure allocations

$$x^{(1)} = (10, 0, 0, 0), \qquad x^{(2)} = (4, 3, 2, 1),$$

and let X_x denote the random variable obtained by choosing a uniformly random index $I \sim \mathrm{Unif}\{1, \ldots, 4\}$ and setting $X_x := x_I$. Then

$$X_{x^{(1)}} \sim \frac{3}{4}\mathbf{1}_{\{0\}} + \frac{1}{4}\mathbf{1}_{\{10\}}, \qquad X_{x^{(2)}} \sim \frac{1}{4}\mathbf{1}_{\{1\}} + \frac{1}{4}\mathbf{1}_{\{2\}} + \frac{1}{4}\mathbf{1}_{\{3\}} + \frac{1}{4}\mathbf{1}_{\{4\}}.$$

Therefore

$$Z \;=\; \frac{1}{2} X_{x^{(1)}} + \frac{1}{2} X_{x^{(2)}},$$

Hence Z is $(10, 4)$-feasible.

Why this matters. The relationship between the three games is:

$$\text{Blotto} \xrightarrow{\text{symmetrise / relabel battlefields}} \text{Colonel Lotto} \xrightarrow{\text{drop feasibility, keep only mean}} \text{General Lotto}.$$

The first arrow preserves the minmax value (and allows us to work with exchangeable play), while the second arrow enlarges the strategy space.

Concretely, General Lotto is a relaxation of Colonel Lotto: in $\Gamma(a, b)$ a player may choose *any* distribution on $\mathbb{Z}_{\geq 0}$ with mean a, whereas in Colonel Lotto the induced one-battlefield marginal must be (A, K)-feasible for $a = A/K$.It must come from randomising over valid K-partitions of A and then selecting a battlefield uniformly at random.

Feasibility problem. Having solved the discrete General Lotto game for all $a \geq b > 0$, the remaining step is to transfer these optimal *marginals* back into an equilibrium construction for Colonel Blotto. To do so, we must determine which General Lotto optimal marginals are (A, K)-feasible and explicitly implement them as distributions over K-partitions of total budget A. This raises the question:

Which General Lotto optimal distributions can be implemented by valid Blotto allocations?

6 Solving symmetric discrete Colonel Blotto via feasibility

Overview. In this project we focus on the symmetric case $A = B$, where Hart's solution follows by combining the General Lotto solutions (Theorems 2–3) in subsection 4.2 with feasibility (Proposition 6), yielding optimal strategies for symmetric Blotto (Theorem 7).

We now

1. State Proposition 6 and Theorem 7
2. Show how the feasibility construction produces an explicit Blotto mixture from the Lotto-optimal marginal
3. Apply the construction to our running instance $A = 5$, $K = 3$

Proposition 6.1 (Proposition 6: Feasibility taken directly from [1]). *Let $A \geq 1$ and $K \geq 2$ be integers.*

i If $A = mK$ where $m \geq 1$ is an integer, then U_O^m is (A, K)-feasible if and only if A and K have the same parity (both are even or both are odd).

ii If $A = mK$ where $m \geq 1$ is an integer, then U_E^m is (A, K)-feasible if and only if A is even.

iii If $A = mK + r$ where $m \geq 0$ and $1 \leq r \leq K - 1$ are integers, then

$$\left(1 - \frac{r}{K}\right) U_E^m + \left(\frac{r}{K}\right) U_O^{m+1}$$

is (A, K)-feasible.

Restatement (integer case). When $A/K = m \in \mathbb{Z}$, parts (i)–(ii) can be restated as follows:

1. If K is even, then $A = mK$ is even, and *both* U_O^m and U_E^m are (A, K)-feasible.
2. If K is odd, then exactly one is feasible:

$$A \text{ even } \Longrightarrow U_E^m \text{ feasible}, \qquad A \text{ odd } \Longrightarrow U_O^m \text{ feasible}.$$

Theorem 6.2 (Theorem 7 taken directly from [1]). *Proposition 6.1 provides optimal strategies for every symmetric Colonel Blotto game $B(A, A; K)$.*

Proof. If $A/K = m$ is an integer, at least one of U_O^m and U_E^m is (A, K)-feasible by Proposition 6.1(i) and (ii), and we apply Theorem 2. Otherwise, $A/K = m + \alpha$ where $\alpha = r/K$ for some $1 \le r \le K - 1$, and then the strategy

$$(1-\alpha)U_E^m \; + \; \alpha\, U_O^{m+1}$$

is (A, K)-feasible by Proposition 6.1(iii); apply Theorem 3. □

Example. Let $A = 8$ and $K = 3$, so

$$\frac{A}{K} = \frac{8}{3} = 2 + \frac{2}{3}, \qquad m = 2, \; r = 2, \; \alpha = \frac{r}{K} = \frac{2}{3}.$$

Proposition 6.1(iii) guarantees that the General-Lotto-optimal marginal

$$(1-\alpha)U_E^2 + \alpha U_O^3 \; = \; \frac{1}{3}U_E^2 + \frac{2}{3}U_O^3$$

is (A, K)-feasible. One explicit feasible implementation is given by the 3×3 matrix

$$S^2 \; = \; \begin{pmatrix} 0 & 3 & 5 \\ 1 & 2 & 5 \\ 1 & 3 & 4 \end{pmatrix},$$

whose rows are 3-partitions of 8. Reading off the induced Blotto mixture,

$$\frac{1}{3}\langle 0,3,5\rangle \; + \; \frac{1}{3}\langle 1,2,5\rangle \; + \; \frac{1}{3}\langle 1,3,4\rangle$$

, where $\langle x \rangle$ is the uniform distribution on all permutations of x.

and then sampling a uniformly random battlefield, the marginal distribution is determined by the multiset of entries of S^2:

$$\text{entries: } \{0,2,4\} \text{ occur once each, and } \{1,3,5\} \text{ occur twice each.}$$

Hence the induced marginal is

$$\frac{1}{9}(\mathbf{1}_0 + \mathbf{1}_2 + \mathbf{1}_4) \; + \; \frac{2}{9}(\mathbf{1}_1 + \mathbf{1}_3 + \mathbf{1}_5) \; = \; \frac{1}{3}U_E^2 + \frac{2}{3}U_O^3,$$

as required.

6.1 Solving our running instance $A = 5$, $K = 3$.

We now apply the preceding framework (General Lotto optimality $\Rightarrow$ feasibility $\Rightarrow$ optimal Blotto strategies) to the concrete game $B(5, 5; 3)$ considered throughout.

Let

$$\frac{A}{K} = \frac{5}{3} = 1 + \frac{2}{3}, \qquad m = 1, \ r = 2, \ \alpha = \frac{r}{K} = \frac{2}{3}.$$

By Proposition 6.1(iii), the General-Lotto-optimal marginal of the form $(1 - \alpha)U_E^m + \alpha U_O^{m+1}$ is $(5, 3)$-feasible, namely

$$(1 - \alpha)U_E^1 + \alpha U_O^2 \ = \ \frac{1}{3}U_E^1 + \frac{2}{3}U_O^2,$$

A feasible implementation is the Blotto mixture

$$\sigma^\star \ = \ \frac{1}{2}\langle 0, 2, 3\rangle \ + \ \frac{1}{2}\langle 1, 1, 3\rangle.$$

Indeed, under $\langle 0, 2, 3\rangle$ the induced marginal is uniform on $\{0, 2, 3\}$, and under $\langle 1, 1, 3\rangle$ it is $(2/3)\mathbf{1}_1 + (1/3)\mathbf{1}_3$. Averaging with weight $1/2$ yields

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 2) = \frac{1}{6}, \qquad \mathbb{P}(X = 1) = \mathbb{P}(X = 3) = \frac{1}{3},$$

so the marginal induced by $\sigma^\star$ is exactly $\frac{1}{3}U_E^1 + \frac{2}{3}U_O^2$.

Since the game is symmetric ($A = B$), its minmax value is

$$v\big(B(5, 5; 3)\big) = 0,$$

and by Theorem 6.2 (using the General Lotto solution, Theorem 3 from Section 4.2), $\sigma^\star$ is minmax-optimal for Player 1. By symmetry the analogous construction gives a minmax-optimal strategy for Player 2. Hence $(\sigma^\star, \sigma^\star)$ is an equilibrium for the symmetric game $B(5, 5; 3)$.

Next Step. This concludes the Hart section: the General Lotto relaxation, together with (A, K)-feasibility, yields the optimal value and explicit symmetric implementable optimal strategies. We now move to the computational half of the project, implementing regret-matching in self-play for the finite Blotto matrix game and evaluating convergence using set-robust metrics.

Remark (What our reduction captures). Hart analyses $B(A, A; K)$ by solving the General Lotto relaxation (with mean $a = b = A/K$) and then restricting to (A, K)-feasible marginals. This recovers the correct value and gives explicit *implementable* optimal strategies for the symmetrised (exchangeable) game, but it does not characterise the entire Nash equilibrium set.

Remark (Beyond the symmetric case). Hart also treats certain non-symmetric regimes. In particular, when the per-battlefield means lie in the same unit interval,

$$m \ < \ \frac{A}{K}, \frac{B}{K} \ < \ m + 1,$$

he derives the exact value and optimal strategies; outside this band he provides explicit upper and lower bounds on the value. We do not pursue these cases here.

7 The finite instance B(5,5;3) as a matrix game

Pure strategies for (A,K)=(5,3). In the labelled Colonel Blotto game, a pure strategy is an allocation vector $x = (x_1, x_2, x_3) \in \mathbb{Z}^3_{\geq 0}$ with $x_1 + x_2 + x_3 = 5$. Thus the common pure strategy set is

$$\mathcal{X} := \Big\{x \in \mathbb{Z}^3_{\geq 0} : \sum_{k=1}^{3} x_k = 5\Big\}, \qquad |\mathcal{X}| = \binom{5+3-1}{3-1} = 21.$$

using stars-and-bars argument.

We fix an ordering $\mathcal{X} = \{x^{(1)}, \dots, x^{(21)}\}$ (used for indexing rows/columns of the payoff matrix) as follows:

1. $(0,0,5)$
2. $(0,1,4)$
3. $(0,2,3)$
4. $(0,3,2)$
5. $(0,4,1)$
6. $(0,5,0)$
7. $(1,0,4)$
8. $(1,1,3)$
9. $(1,2,2)$
10. $(1,3,1)$
11. $(1,4,0)$
12. $(2,0,3)$
13. $(2,1,2)$
14. $(2,2,1)$
15. $(2,3,0)$
16. $(3,0,2)$
17. $(3,1,1)$
18. $(3,2,0)$
19. $(4,0,1)$
20. $(4,1,0)$
21. $(5,0,0)$

Payoff and matrix construction. Recall definition 1.1, given $x, y \in \mathcal{X}$, define the per-battlefield outcome by

$$\operatorname{sgn}(x_k - y_k) = \begin{cases} +1, & x_k > y_k, \\ 0, & x_k = y_k, \\ -1, & x_k < y_k, \end{cases}$$

and the normalised Blotto payoff to Player 1 by

$$u(x,y) \;:=\; \frac{1}{3}\sum_{k=1}^{3} \operatorname{sgn}(x_k - y_k) \;\in\; \{-1, -\tfrac{2}{3}, -\tfrac{1}{3}, 0, \tfrac{1}{3}, \tfrac{2}{3}, 1\}.$$

We now form the zero-sum matrix game with payoff matrix $M \in \mathbb{R}^{21\times 21}$ defined by

$$M_{ij} \;:=\; u\big(x^{(i)}, x^{(j)}\big), \qquad i, j \in \{1, \dots, 21\}.$$

Since $u(x,y) = -u(y,x)$, the matrix M is skew-symmetric: $M = -M^\top$, and in particular the value of the symmetric game is 0 which matches Hart's result.

(See Appendix B.1 for an example of a subgame payoff matrix)

Lemma 7.1. *Value of the symmetric game is 0.*

Proof. Let $M \in \mathbb{R}^{n\times n}$ be the payoff matrix of a finite two-player zero-sum game, and let

$$v \;=\; \max_{x\in\Delta_n} \min_{y\in\Delta_n} x^\top M y \;=\; \min_{y\in\Delta_n} \max_{x\in\Delta_n} x^\top M y$$

denote its value (by the minmax theorem 1.1). As M is skew-symmetric, $M = -M^\top$. Then for any $x, y \in \Delta_n$,

$$x^\top M y \;=\; (x^\top M y)^\top \;=\; y^\top M^\top x \;=\; -\,y^\top M x.$$

Therefore,

$$\begin{aligned} v &= \max_{x\in\Delta_n}\min_{y\in\Delta_n} x^\top My = \max_{x\in\Delta_n}\min_{y\in\Delta_n}\left(-y^\top Mx\right) \\ &= -\min_{x\in\Delta_n}\max_{y\in\Delta_n} y^\top Mx = -\min_{y\in\Delta_n}\max_{x\in\Delta_n} x^\top My = -v. \end{aligned}$$

Hence $v = -v$, so $v = 0$. □

(See Appendix B.2 for an example of computing the expected payoff.)

8 Regret learning

Overview. Regret learning studies repeated decision-making in games where a player updates their choices using feedback from past rounds. One of the key notion is *external regret*: over time, the learner should do nearly as well as the best *single fixed action* they could have committed to in hindsight. In this project we implement *regret-matching* in the practical "trainer" form outlined by Neller–Lanctot [2].

Set-up. Fix a player with a finite action set $\mathcal{A} = \{1, \ldots, m\}$. At rounds $t = 1, 2, \ldots$ the player selects an action $a_t \in \mathcal{A}$, the opponent selects $b_t \in \mathcal{B}$, and the player receives payoff $u(a_t, b_t) \in \mathbb{R}$. We have payoff matrix $M \in \mathbb{R}^{m\times m}$ for Player 1 with $u(i, j) = M_{ij}$.

External regret. External regret compares the realised payoff sequence to the payoff that would have been obtained by always playing a *fixed* action $i \in \mathcal{A}$.

Definition 8.1 (cumulative regret for action i up to time T and average external regret)**.**

$$R_i(T) \;:=\; \sum_{t=1}^{T}\big(u(i, b_t) - u(a_t, b_t)\big), \qquad R_i^+(T) := \max\{R_i(T), 0\}. \tag{1}$$

The average external regret is

$$\bar{R}(T) \;:=\; \frac{1}{T}\max_{i\in\mathcal{A}} R_i^+(T). \tag{2}$$

A learning rule is ***no-external-regret*** if $\bar{R}(T) \to 0$ as $T \to \infty$.

Regret-matching. Given cumulative regrets $R(T) = (R_1(T), \ldots, R_m(T))$, regret-matching chooses the next mixed action $p_{T+1} \in \Delta(\mathcal{A})$ by normalising the *positive* regrets:

$$p_{T+1}(i) \;=\; \begin{cases} \dfrac{R_i^+(T)}{\sum_{k\in\mathcal{A}} R_k^+(T)}, & \text{if } \sum_{k\in\mathcal{A}} R_k^+(T) > 0, \\ \dfrac{1}{m}, & \text{otherwise.} \end{cases} \tag{3}$$

Interpretation: actions with larger positive regret are ones we "wish" we had played more often, so we increase their probability. If all regrets are non-positive, we return to uniform play.

8.1 Neller–Lanctot regret matching algorithm [2].

Let $R^{(t)} \in \mathbb{R}^m$ denote the cumulative regret after t iterations and let $S^{(t)} \in \mathbb{R}^m$ be the running sum of mixed strategies.

Initialise $R^{(0)} = \mathbf{0}$ and $S^{(0)} = \mathbf{0}$.

Remark. We start off with uniform mixed actions for both players at t=1

For $t = 1, 2, \ldots, T$:

1. **Regret-matched strategy:** compute p_t using the rule above.
2. **Accumulate average:** update $S^{(t)} \leftarrow S^{(t-1)} + p_t$.
3. **Sample play:** sample realised actions $a_t \sim p_t$ and $b_t \sim q_t$.
4. **Regret update:** update regrets using the realised opponent action b_t:

$$R_i^{(t)} = R_i^{(t-1)} + \big(u(i, b_t) - u(a_t, b_t)\big), \qquad \forall i \in \mathcal{A}. \tag{4}$$

   equivalently with $u(i, b_t) = M_{i\, b_t}$,

$$R^{(t)} = R^{(t-1)} + \underbrace{M_{:,b_t}}_{\text{the } b_t\text{-th column of } M} - M_{a_t,b_t}\,\mathbf{1}.$$

Finally, the reported strategy is the time-average of the mixed strategies:

$$\bar{p}_T := \frac{1}{T}\sum_{t=1}^{T} p_t = \frac{1}{T} S^{(T)}. \tag{5}$$

(Analogous updates are applied to Player 2. Recall, in the zero-sum Blotto instance Player 2's payoff matrix is $-M$.)

Why time-averaging is the output. Individual iterates p_t can fluctuate substantially due under sampling), whereas the average $\bar{p}_T$ stabilises as T increases. Therefore we evaluate $\bar{p}_T$ (and similarly $\bar{q}_T$) rather than the final iterate p_T.

8.2 Convergence of regret learning and the time-averaged mixed strategy.

This subsection uses ideas presented in Prediction, Learning, and Games [3] and Regret Matching+: (In)Stability and Fast Convergence in Games [5].

General-sum interpretation. If each player achieves vanishing external regret (Hannan consistency), then the empirical joint distribution of play $\bar{P}_T$ approaches the Hannan set (the set of coarse correlated equilibria).

Zero-sum interpretation. In two-player zero-sum games, if both players achieve vanishing external regret, then the average payoff converges to the **game value** (Theorem 7.2 and Corollary 7.1), and the product of the marginal empirical distributions $(\bar{p}_T, \bar{q}_T)$ converges to the **Nash equilibrium set** (Remark 7.4).

(See Appendix A.2 for Theorem 7.2, Corollary 7.1 and remark 7.4.)

Remark (Using $\bar{p}_T$ instead of $\hat{p}_T$). Remark 7.4 is stated for the empirical action frequencies $(\hat{p}_T, \hat{q}_T)$, where $\hat{p}_T(i) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{I_t = i\}$ and $\hat{q}_T(j) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{J_t = j\}$. In our experiments we track the time-averaged mixed strategies $\bar{p}_T = \frac{1}{T}\sum_{t=1}^{T} p_t$ and $\bar{q}_T = \frac{1}{T}\sum_{t=1}^{T} q_t$. Since $I_t \sim p_t$ and $J_t \sim q_t$, for each i the sequence $\mathbf{1}\{I_t = i\} - p_t(i)$ is a bounded martingale difference, hence $\hat{p}_T(i) - \bar{p}_T(i) \to 0$ and $\hat{q}_T(j) - \bar{q}_T(j) \to 0$ as $T \to \infty$ (almost surely, and in particular in probability). Therefore $\|\hat{p}_T - \bar{p}_T\|_1 \to 0$ and $\|\hat{q}_T - \bar{q}_T\|_1 \to 0$, so any continuity-based conclusion (in exploitability) transfers from $(\hat{p}_T, \hat{q}_T)$ to $(\bar{p}_T, \bar{q}_T)$. Furthermore, Remark 7.4 mentions that this does *not* imply convergence of the joint empirical distribution of play.

Convergence rate of regret learning. Recall the (external) regret components $R_i(T) = \sum_{t=1}^{T}\big(u(i, b_t) - u(a_t, b_t)\big)$ and the per-round regret

$$\bar{R}(T) \;:=\; \frac{1}{T}\max_{i\in\mathcal{A}} R_i^+(T).$$

Regret Matching (RM) is a regret minimiser on the simplex with worst-case cumulative external regret $\max_{i\in\mathcal{A}} R_i(T) = O(\sqrt{T})$ (up to problem-dependent constants), and therefore its per-round regret decays as

$$\bar{R}(T) = O\left(\frac{1}{\sqrt{T}}\right).$$

This $\sqrt{T}$-type guarantee for regret minimisers (including RM and its variants) is discussed in Regret Matching+: (In)Stability and Fast Convergence in Games [5].

9 Verification Methods: Exploitability and Hart-marginal distance

Overview. This section introduces two verification tools. First, we use *exploitability* to verify that the time-averaged strategies produced by regret matching are approximately Nash in the finite zero-sum game. Second, we compare Hart's prediction to the learned play by measuring the distance between Hart's theoretical random-battlefield marginal and the marginal induced by the symmetrised regret-matching strategy.

9.1 Exploitability

The following definition is adapted from Approximate Exploitability: Learning a Best Response [6] Section 2.2.

Definition 9.1 (Player exploitabilities, NashConv, exploitability)**.** For any pair $(p, q) \in \Delta_n \times \Delta_n$ define the *player-specific deviation incentives*

$$\varepsilon_1(p,q) := BR_1(q) - u(p,q) \;=\; \max_i (Mq)_i - p^\top Mq,$$
$$\varepsilon_2(p,q) := u(p,q) - BR_2(p) \;=\; p^\top Mq - \min_j (p^\top M)_j.$$

Define the *NashConv* of (p, q) by

$$\mathrm{NashConv}(p,q) \;:=\; \varepsilon_1(p,q) + \varepsilon_2(p,q),$$

and the (two-player) *exploitability* by

$$\mathrm{Exploit}(p,q) \;:=\; \frac{1}{2}\mathrm{NashConv}(p,q) \;=\; \frac{\varepsilon_1(p,q) + \varepsilon_2(p,q)}{2}.$$

(In an n-player setting one uses $\mathrm{Exploitability} = \mathrm{NashConv}/n$.)

In particular, (p, q) is a Nash equilibrium iff $\varepsilon_1(p, q) = \varepsilon_2(p, q) = 0$, equivalently iff $\mathrm{NashConv}(p, q) = 0$ and hence $\mathrm{Exploit}(p, q) = 0$.

Interpretation.

- $\varepsilon_1(p, q)$ is the amount by which Player 1 can *increase* payoff by deviating to a best response to q.
- $\varepsilon_2(p, q)$ is the amount by which Player 2 can *decrease Player 1's payoff* by deviating to a best response to p (equivalently, the amount by which Player 2 can increase their own payoff in the zero-sum game).
- $\mathrm{NashConv}(p, q)$ is the *total* one-shot improvement available to the two players.
- $\mathrm{Exploit}(p, q)$ is the *average* deviation incentive per player.

Convergence rate for exploitability for time-averaged mixed strategies. We first show that, in a matrix game, each player's deviation incentive at the time-averaged strategies is exactly its per-round external regret.

Let $u(p, q) = p^\top M q$ and define the time-averaged strategies

$$\bar{p}_T = \frac{1}{T}\sum_{t=1}^{T} p_t, \qquad \bar{q}_T = \frac{1}{T}\sum_{t=1}^{T} q_t.$$

The row player's per-round external regret is

$$\bar{R}_1(T) := \frac{1}{T}\Big(\max_{i\in\mathcal{A}} \sum_{t=1}^{T} u(e_i, q_t) - \sum_{t=1}^{T} u(p_t, q_t)\Big).$$

By bilinearity of u,

$$\frac{1}{T}\sum_{t=1}^{T} u(e_i, q_t) = u(e_i, \bar{q}_T) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T} u(p_t, q_t) = u(\bar{p}_T, \bar{q}_T),$$

hence

$$\bar{R}_1(T) = \max_{i\in\mathcal{A}} u(e_i, \bar{q}_T) - u(\bar{p}_T, \bar{q}_T) =: \varepsilon_1(\bar{p}_T, \bar{q}_T).$$

Similarly, the column player's per-round external regret equals its deviation incentive

$$\bar{R}_2(T) = \ u(\bar{p}_T, \bar{q}_T) - \min_{j\in\mathcal{A}} u(\bar{p}_T, e_j) =: \varepsilon_2(\bar{p}_T, \bar{q}_T).$$

Therefore,

$$\mathrm{NashConv}(\bar{p}_T, \bar{q}_T) = \varepsilon_1(\bar{p}_T, \bar{q}_T) + \varepsilon_2(\bar{p}_T, \bar{q}_T) = \bar{R}_1(T) + \bar{R}_2(T), \qquad \mathrm{Exploit}(\bar{p}_T, \bar{q}_T) = \tfrac{1}{2}\,\mathrm{NashConv}(\bar{p}_T, \bar{q}_T).$$

From the previous subsection 8.2, RM guarantees $\bar{R}_1(T), \bar{R}_2(T) = O(1/\sqrt{T})$, and hence

$$\mathrm{NashConv}(\bar{p}_T, \bar{q}_T) = O\left(\frac{1}{\sqrt{T}}\right),$$

$$\mathrm{Exploit}(\bar{p}_T, \bar{q}_T) = O\left(\frac{1}{\sqrt{T}}\right).$$

9.2 Comparing Hart structure via total variation distance

Recall from section 6, Hart's prediction for $B(5,5;3)$ is stated in terms of the *random-battlefield marginal* $P(X = t)$. We therefore compare the marginal induced by the symmetrised regret-matching strategy to Hart's theoretical marginal using **total variation distance**.

Inducing a marginal distribution from the symmetrised regret-matching strategy. Let $\bar{p}_T \in \Delta(\mathcal{X})$ denote the time-averaged mixed strategy over pure allocations $x = (x_1, \ldots, x_K) \in \mathcal{X}$ with $\sum_{k=1}^{K} x_k = A$, and let $\sigma(\bar{p}_T)$ be its symmetrisation over permutation classes.

From any symmetrised mixed strategy $\mu \in \Delta(\mathcal{X})$ we induce the *random-battlefield marginal* P_μ on troop counts $t \in \{0, 1, \ldots, A\}$ by sampling an allocation $x \sim \mu$ and then sampling a battlefield index I uniformly from $\{1, \ldots, K\}$, and returning $t = x_I$. Equivalently,

$$P_\mu(t) \;=\; \frac{1}{K} \sum_{x \in \mathcal{X}} \mu(x) \left|\{\, i \in \{1, \ldots, K\} : x_i = t \,\}\right|, \qquad t = 0, 1, \ldots, A.$$

(See Appendix B.3 for an example for (A=5, K=3).)

We then compare the resulting induced marginal $P_{\sigma(\bar{p}_T)}$ to Hart's target marginal using d_{TV}.

Definition (total variation distance). For two probability distributions P, Q on the finite set $\{0, 1, \ldots, A\}$, define

$$d_{\mathrm{TV}}(P, Q) \;:=\; \frac{1}{2} \sum_{t=0}^{A} \left|P(t) - Q(t)\right|.$$

Interpretation. $d_{\mathrm{TV}}(P, Q) \in [0, 1]$, with $d_{\mathrm{TV}}(P, Q) = 0$ iff $P = Q$. Smaller values mean the two marginals are closer.

10 Implementation

Overview. This section describes the full implementation outline for our running instance $B(5,5;3)$: constructing the finite Blotto game (action set and payoff matrix), running regret matching from Neller–Lanctot [2], and computing the averaged and symmetrised strategies used for evaluation. We also provide pseudocode for regret matching and the symmetrisation method for Hart comparison.

1. **Action set construction (Blotto allocations).**
   - Fix the running instance $(A, K) = (5, 3)$.
   - Enumerate

     $$\mathcal{X} = \Big\{ x \in \mathbb{Z}_{\geq 0}^{K} : \sum_{i=1}^{K} x_i = A \Big\}, \qquad |\mathcal{X}| = \binom{A + K - 1}{K - 1} = 21,$$

     and store an ordered list `actions` $= \{x^{(1)}, \ldots, x^{(|\mathcal{X}|)}\}$ together with an index map `idx` $: \mathcal{X} \to \{1, \ldots, |\mathcal{X}|\}$.
2. **Payoff matrix.**

- Implement the stage payoff $u(x, y)$ for $x, y \in \mathcal{X}$.
- Form the payoff matrix $M \in \mathbb{R}^{|\mathcal{X}|\times|\mathcal{X}|}$ by

$$M_{ij} = u\big(x^{(i)}, x^{(j)}\big).$$

3. **Regret matching (sampled-action self-play).**

Algorithm 1: Regret Matching (sampled-action self-play, adapted from Neller–Lanctot [2])

Input: Payoff matrix $M \in \mathbb{R}^{n\times n}$ (payoff to Player 1), iterations T, optional seed
Output: Averaged strategies $\bar{p}_T, \bar{q}_T \in \Delta_n$ and logs $\{u_t, v_t\}_{t=1}^T$

Initialise:
$R^A \leftarrow 0 \in \mathbb{R}^n, \quad R^B \leftarrow 0 \in \mathbb{R}^n$
$\mathsf{sumP} \leftarrow 0 \in \mathbb{R}^n, \quad \mathsf{sumQ} \leftarrow 0 \in \mathbb{R}^n$

for $t = 1$ **to** T **do**

```
    // 1) Strategy update from positive regrets (fallback to uniform)
    p_t ← NormalizePos(R^A)
    q_t ← NormalizePos(R^B)
    // 2) Diagnostic: expected payoff under mixed strategies
    v_t ← p_t^T M q_t
    // 3) Sample realised pure actions
    i_t ~ Sample(p_t)                                  // Player 1 plays row i_t
    j_t ~ Sample(q_t)                                  // Player 2 plays column j_t
    u_t ← M_{i_t j_t}                                  // realised payoff to Player 1
    // 4) Realised external-regret updates
    R^A ← R^A + (M_{:,j_t} − u_t 1)                    // M_{:,j_t} is column j_t
    R^B ← R^B + (u_t 1 − M_{i_t,:})                    // M_{i_t,:} is row i_t
    // 5) Averaging of mixed strategies
    sumP ← sumP + p_t
    sumQ ← sumQ + q_t
```

$\bar{p}_T \leftarrow \mathsf{sumP}/T, \quad \bar{q}_T \leftarrow \mathsf{sumQ}/T$
return $(\bar{p}_T, \bar{q}_T, \{u_t, v_t\}_{t=1}^T)$

Algorithm 2: NormalizePos(R)

Input: $R \in \mathbb{R}^n$
Output: $p \in \Delta_n$
$z \leftarrow \max(R, 0)$ // elementwise positive part
if $\sum_{k=1}^n z_k > 0$ **then** $p \leftarrow z / \sum_{k=1}^n z_k$
else
  $p \leftarrow \frac{1}{n}\mathbf{1}$
return p

4. **Symmetrisation (for Hart comparison).**

Why we symmetrise learned strategies. Hart's prediction is stated for the symmetrised formulation (equivalently, Colonel Lotto; see Lemma 3.1). Since regret-matching outputs a mixed strategy over labelled allocations, we symmetrise the time-average $\bar{p}_T$ by averaging over permutation classes, obtaining $\sigma(\bar{p}_T)$ whose induced marginal is directly comparable to Hart's.

In our experiments we apply the symmetrisation operator only at evaluation time, to the final time-averaged strategies $\bar{p}_T$ and $\bar{q}_T$.

Remark (Why symmetrise only at evaluation). We apply σ only to the final time-averaged strategies $\bar{p}_T, \bar{q}_T$. Symmetrising at every iteration would modify the regret-matching dynamics (effectively projecting each iterate onto the symmetric subspace), whereas evaluating at time T is a harmless relabeling-invariant summary used only for comparison with Hart/Lotto.

Algorithm 3: Symmetrisation of a mixed strategy via permutation classes

---

Input: Action list `actions` $= \{x^{(1)}, \dots, x^{(n)}\} \subset \mathcal{X}$, mixed strategy $p \in \Delta_n$
Output: $p^{\text{sym}} = \sigma(p) \in \Delta_n$

Initialise:
mass $\leftarrow$ empty map from sorted-patterns to $\mathbb{R}$
count $\leftarrow$ empty map from sorted-patterns to $\mathbb{N}$
$p^{\text{sym}} \leftarrow 0 \in \mathbb{R}^n$

`// Pass 1: accumulate permutation-class masses and class sizes`
for $i = 1$ **to** n **do**
  $c \leftarrow \text{sort}(x^{(i)})$
  mass$[c]$ $\leftarrow$ mass$[c]$ $+ p_i$
  count$[c]$ $\leftarrow$ count$[c]$ $+ 1$

`// Pass 2: redistribute each class mass uniformly across its members`
for $i = 1$ **to** n **do**
  $c \leftarrow \text{sort}(x^{(i)})$
  $p_i^{\text{sym}} \leftarrow$ mass$[c]$/count$[c]$
return p^{sym}

Algorithm 4: Compute symmetrised averages for RM outputs

Input: $\bar{p}_T, \bar{q}_T \in \Delta_n$, action list `actions`
Output: $p_T^{\text{sym}}, q_T^{\text{sym}} \in \Delta_n$
$p_T^{\text{sym}} \leftarrow \textsc{Symmetrize}(\texttt{actions}, \bar{p}_T)$;
$q_T^{\text{sym}} \leftarrow \textsc{Symmetrize}(\texttt{actions}, \bar{q}_T)$;
return $(p_T^{\text{sym}}, q_T^{\text{sym}})$;

---

5. **Verification** (see Section 9)

11 Result

Overview. I aim to answer three main empirical questions on the $(5, 3)$ Colonel Blotto instance:

i Does the game have value 0 as predicted by Hart and skew-symmetry argument ii Does regret-matching result approach the *set* of Nash equilibria, and does its convergence rate match the $O(T^{-1/2})$ behaviour suggested by regret-minimisation theory

iii How closely do the learned symmetrised strategies match with Hart's predicted battlefield marginal.

To address these questions, I implemented the pipeline described in Section 10 in Python and ran the main experiment with parameters $T = 10{,}000$ and `nseeds`$= 50$ (number of seeds), generating the figures and summary statistics reported below.

For full implementation details and reproducibility instructions, see the accompanying GitHub repository:

`https://github.com/Scott-Yap/Dynamic-of-learning-and-iterated-games-project-3`.

11.1 Q1: Verifying that the game value is 0

Numerical estimate of the value. Our run with $T = 10{,}000$ iterations outputs

$$v(\bar{p}_T, \bar{q}_T) = \bar{p}_T^\top M \, \bar{q}_T \approx -1.00515 \times 10^{-4}, \qquad v\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big) \approx -9.86701 \times 10^{-5},$$

both of which are very close to 0. This matches the theoretical argument (skew-symmetry) and Hart's prediction that the value of the symmetric game is zero.

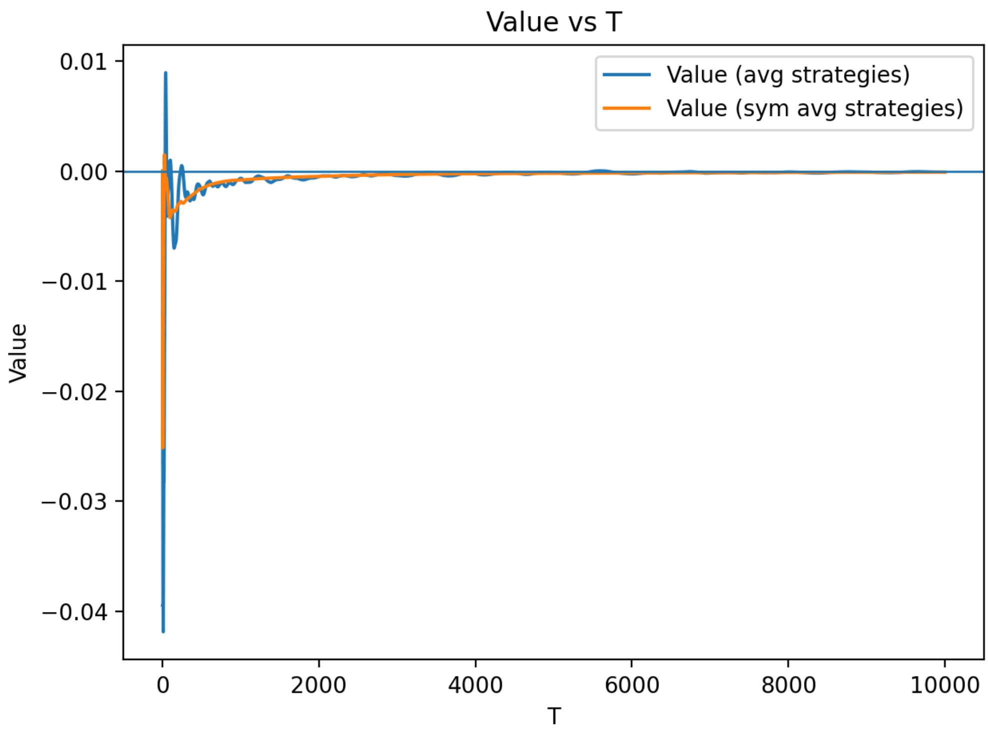

Figure 2 Estimated value as a function of T (main seed). Both $v(\bar{p}_T, \bar{q}_T)$ and $v(\sigma(\bar{p}_T), \sigma(\bar{q}_T))$ stabilise near 0, consistent with the theoretical value.

Interpretation. Figure 2 shows the value estimate approaching 0 as T increases. The larger fluctuations for small T (roughly $T \leq 1000$) are due to the sampled-action updates, which have higher variance early on. As regrets accumulate and the time averages stabilise, both value estimates settle close to 0.

11.2 Q2: Nash verification and convergence rate

11.2.1 Q2(i) Exploitability decreases with T (approach to the Nash set)

Numerical result. For the main run at $T = 10{,}000$ we obtain

$$\mathrm{Exploit}(\bar{p}_T, \bar{q}_T) = 4.82443 \times 10^{-3}, \qquad \mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big) = 1.15327 \times 10^{-4}.$$

Across 50 seeds, the exploitability (non-symmetrised averages) has

$$\mathrm{median}\big(\mathrm{Exploit}(\bar{p}_T, \bar{q}_T)\big) = 4.67566 \times 10^{-3}, \qquad \mathrm{IQR} = [3.95538,\ 5.29212] \times 10^{-3},$$

with range $[2.20274 \times 10^{-3},\ 7.34457 \times 10^{-3}]$. Using symmetrised averages, the exploitability has

$$\mathrm{median}\Big(\mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big)\Big) = 1.43677 \times 10^{-4}, \qquad \mathrm{IQR} = [1.08503,\ 1.85095] \times 10^{-4},$$

with range $[6.41080 \times 10^{-5},\ 6.25998 \times 10^{-4}]$. Overall, these exploitability values are small, showing that the time-averaged mixed strategies approaches the Nash equilibrium set. Moreover, the symmetrised averages have substantially lower exploitability.

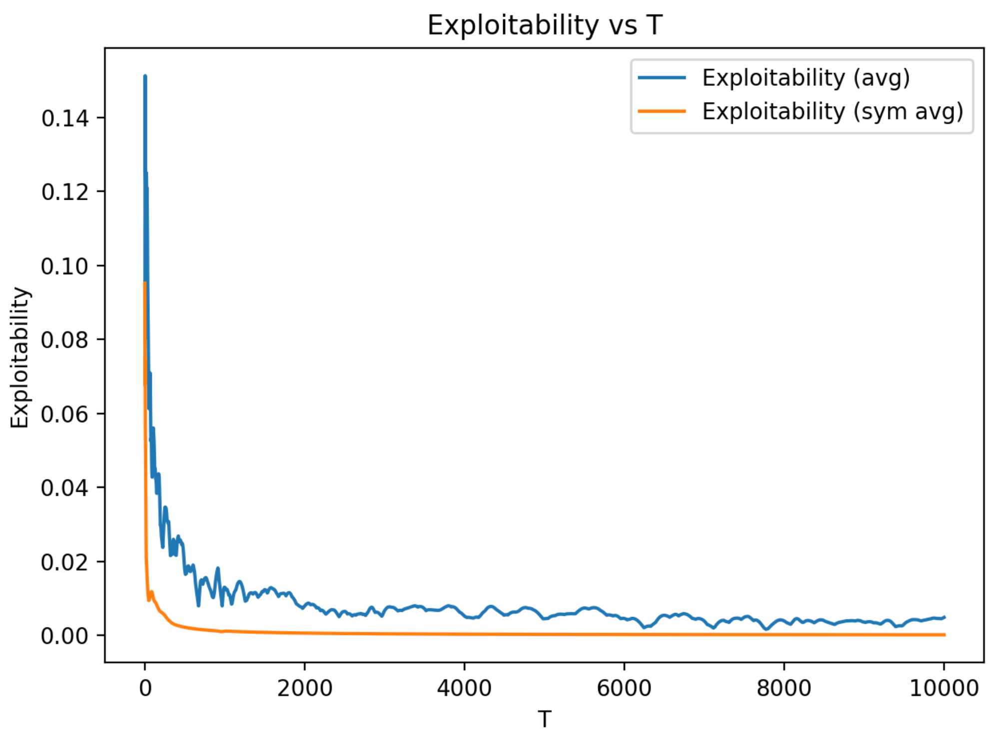

Figure 3 Exploitability vs. T for the averaged strategies (blue) and for the symmetrised averaged strategies (orange).

Interpretation. Figure 3 shows that the exploitability of the time-averaged strategies $\mathrm{Exploit}(\bar{p}_T, \bar{q}_T)$ decreases towards 0 as T increases.The *symmetrised post-processed* averages $\mathrm{Exploit}(\sigma(\bar{p}_T), \sigma(\bar{q}_T))$ are consistently smaller at the same horizon T, we treat this as an empirical effect rather than a general guarantee.

Why does symmetrisation help. Since the Blotto payoff is invariant to battlefield relabelling, we can w.l.o.g. restrict attention to exchangeable strategies. Applying σ as a *post-processing step* enforces permutation-invariance on $\bar{p}_T, \bar{q}_T$. In this instance it empirically lowers exploitability, plausibly by removing finite-T label-asymmetries. However, symmetrisation does not come with a general theorem that it must improve exploitability.

11.2.2 Q2(ii) Convergence rate: $O(T^{-1/2})$

Numerical Result. To test the convergence rate numerically, we plot rescaled exploitability curves across 50 seeds. For the unsymmetrised averages we consider $\sqrt{T}\,\mathrm{Exploit}(\bar{p}_T, \bar{q}_T)$, and for the symmetrised averages we consider $T\,\mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big)$. Dropping the early transient (small T) and summarising only the stabilised tail, the rescaled constants (median with IQR *across seeds*) are

$$\mathrm{median}\big(\sqrt{T}\,\mathrm{Exploit}(\bar{p}_T, \bar{q}_T)\big) = 0.493, \qquad \mathrm{IQR} = [0.447,\ 0.537],$$

and

$$\mathrm{median}\Big(T\,\mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big)\Big) = 1.44, \qquad \mathrm{IQR} = [1.09,\ 1.85].$$

We will use these values to aid our interpretation of the plot.

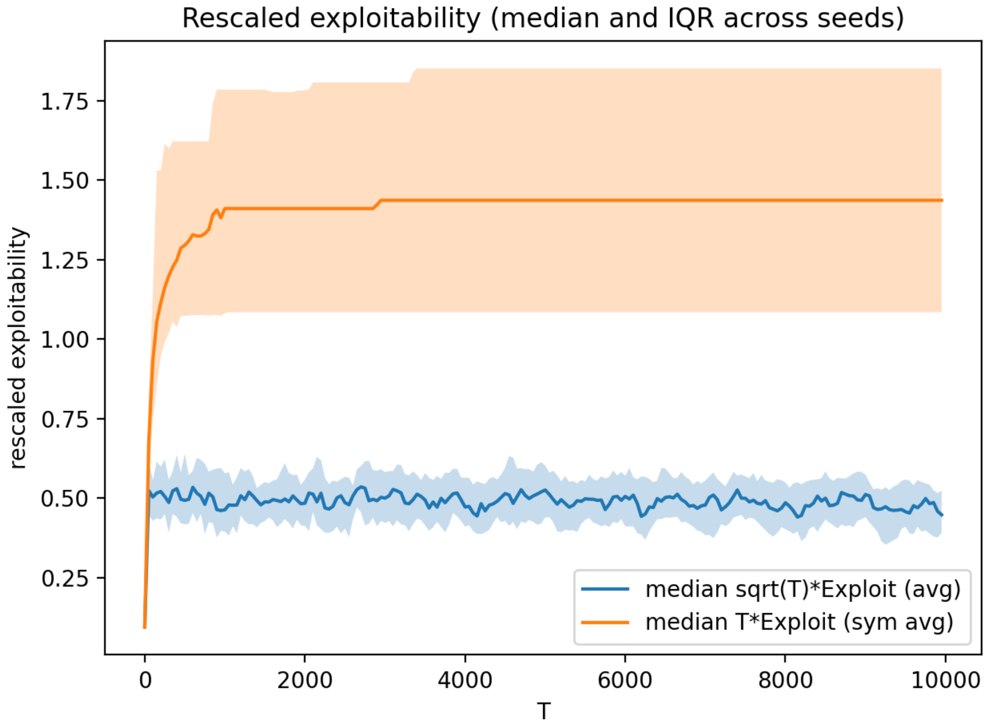

Figure 4 Rescaled Nash-gap (median $\pm$ IQR across seeds). Blue: $\sqrt{T}\,\varepsilon(\bar{p}_T, \bar{q}_T)$. Orange: $T\,\varepsilon(\sigma(\bar{p}_T), \sigma(\bar{q}_T))$.

Interpretation. Figure 4 shows that the rescaled curve $\sqrt{T}\,\mathrm{Exploit}(\bar{p}_T, \bar{q}_T)$ is relatively stable for large T, which is consistent with the theoretical $O(T^{-1/2})$ convergence rate in subsection 9.1.

For the symmetrised averages, the curve $T\,\mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big)$ is empirically close to flat for large T, which is suggestive of an approximately $\frac{1}{T}$ decay for this instance, although we do not claim a general theoretical guarantee for this behaviour.

Finally, the shaded band is wider for the symmetrised curve because the across-seed constants vary much more after symmetrisation: applying σ removes labelling noise, so the remaining differences come from how each seed distributes mass across permutation classes, and the max/min in exploitability amplifies these small differences.

11.3 Q3: Comparison with Hart's marginal prediction

11.3.1 Q3(i) Hart's *strategy* lies in the Nash set

Recall from Subsection 6.1 that for the running instance $B(5, 5; 3)$ Hart's General-Lotto-optimal *random-battlefield marginal* is

$$P^\star(0) = P^\star(2) = \tfrac{1}{6}, \qquad P^\star(1) = P^\star(3) = \tfrac{1}{3}, \qquad P^\star(4) = P^\star(5) = 0.$$

To verify Nash-consistency in the *finite* Blotto matrix game, we do not test the marginal directly, but instead test a mixed strategy that implements it. Using the feasible implementation from Subsection 6.1, we take the symmetric mixed strategy

$$\sigma^\star \;=\; \tfrac{1}{2}\,\langle 0, 2, 3\rangle \;+\; \tfrac{1}{2}\,\langle 1, 1, 3\rangle,$$

and evaluate the Nash inequalities via exploitability at $(\sigma^\star, \sigma^\star)$ in our payoff matrix M:

$$Exploit(\sigma^\star, \sigma^\star) \approx 0.0, \qquad \varepsilon_1 \approx 0.0, \qquad \varepsilon_2 \approx 0.0.$$

Hence neither player can gain by a unilateral deviation, so $(\sigma^\star, \sigma^\star)$ is an exact Nash equilibrium of $B(5, 5; 3)$. In particular, Hart's target marginal $P^\star$ is attained by an equilibrium strategy profile, so it is consistent with the Nash set.

11.3.2 Q3(ii) Marginal comparison at the final iterate

Interpretation. Figure 5 compares the marginal distribution of troops on a uniformly random battlefield induced by the symmetrised learned strategy $\sigma(\bar{p}_T)$ with Hart's target marginal. Both place essentially all mass on $\{0, 1, 2, 3\}$ and assign similar probability to 3, but there are visible discrepancies at $0, 1, 2$. At finite T, such differences are expected and can also arise because regret-matching may converge to a different equilibrium mixture within the Nash set than the particular marginal highlighted by Hart.

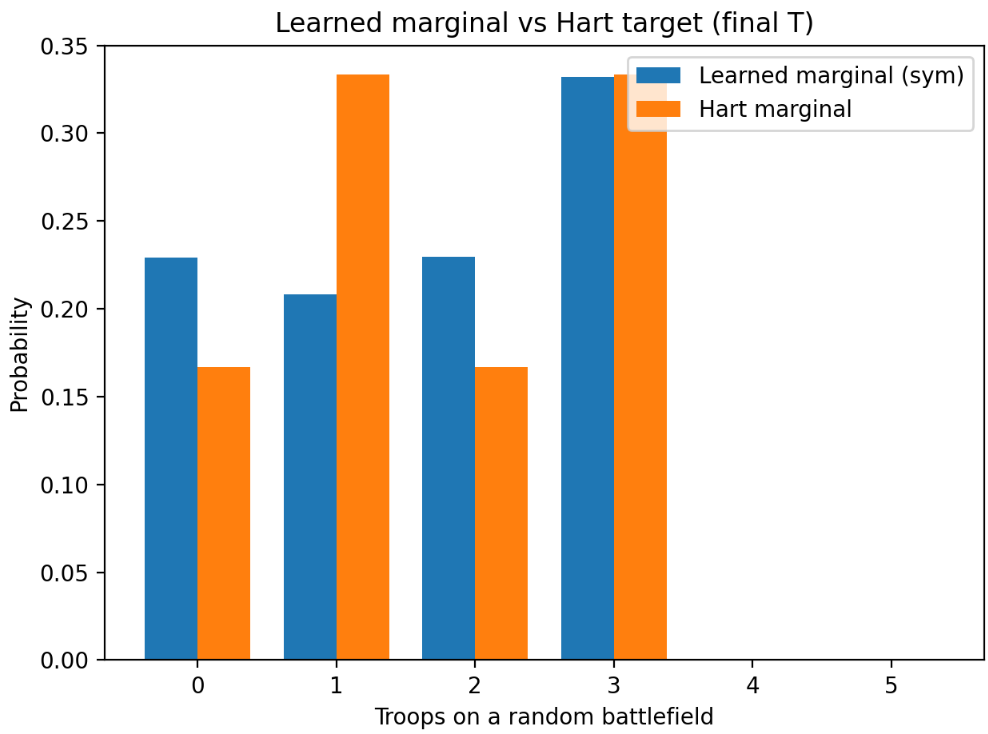

Figure 5 Learned marginal (from $\sigma(\bar{p}_T)$ at the final T) versus Hart's target marginal.

11.3.3 Q3(iii) Total variation distance and seed-to-seed variability

Numerical result. For the main run (seed 0) at $T = 10{,}000$ we obtain

$$d_{\mathrm{TV}}\left(\pi_T, \pi^{\mathrm{Hart}}\right) = 0.126458.$$

Across 50 seeds (using symmetrised averages), the distance to Hart has median

$$\mathrm{median}\left(d_{\mathrm{TV}}(\pi_T, \pi^{\mathrm{Hart}})\right) = 0.116364, \qquad \mathrm{IQR} = [0.100406,\ 0.127012],$$

with range $[0.0346792,\ 0.186287]$. The distance is therefore non-negligible, showing that the learned marginal is qualitatively similar to Hart's target but not identical in general.

Remark. The minimum observed distance across seeds is 0.0346792, indicating that some trajectories come much closer to Hart's marginal than others.

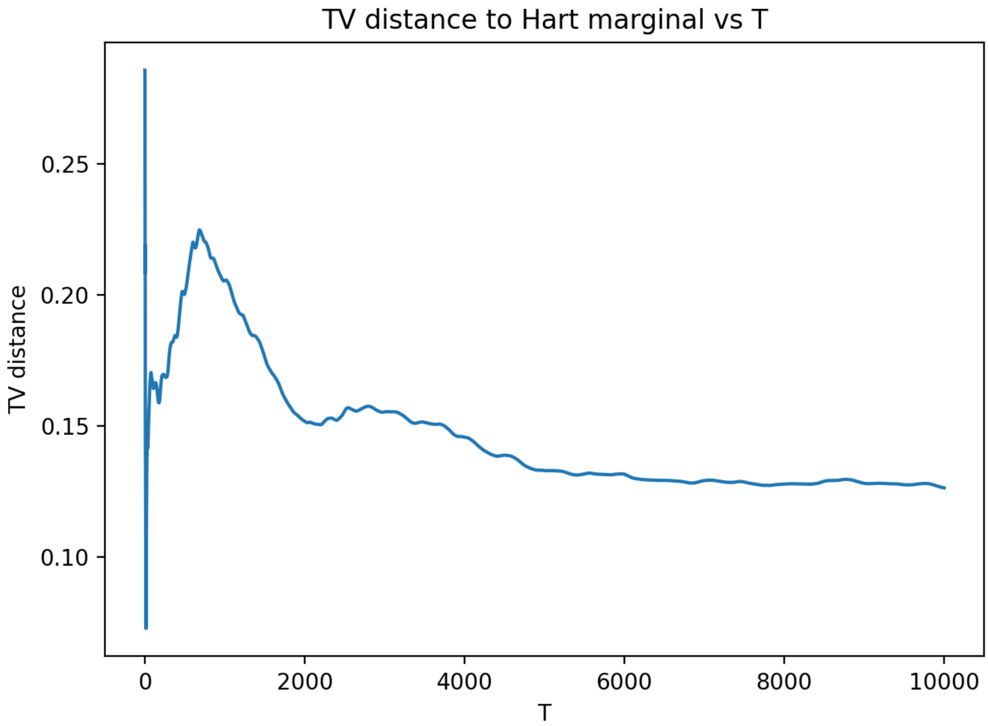

Figure 6 Total variation distance $d_{\mathrm{TV}}(\pi_T, \pi^{\mathrm{Hart}})$ as a function of T (main seed).

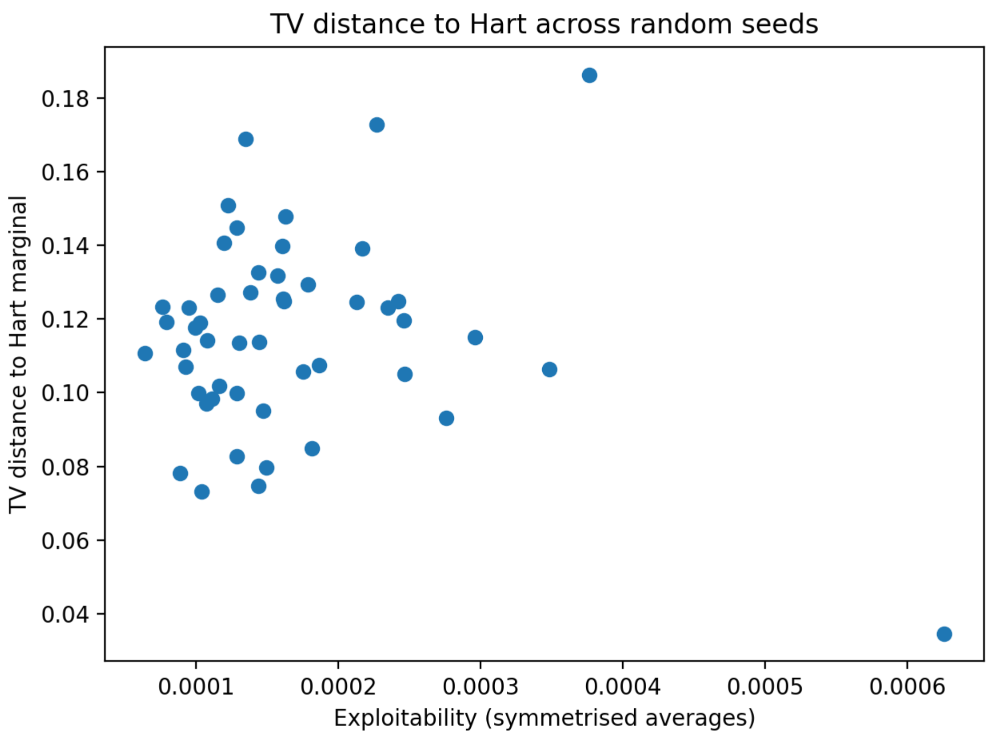

Figure 7 Across seeds: $d_{\mathrm{TV}}(\pi_T, \pi^{\mathrm{Hart}})$ versus $\mathrm{Exploit}\big(\sigma(\bar{p}_T), \sigma(\bar{q}_T)\big)$.

Interpretation. Figure 6 shows $d_{\mathrm{TV}}(\pi_T, \pi^{\mathrm{Hart}})$ decaying from an early transient and then levelling off near 0.12–0.13 for this run.

Figure 7 shows substantial spread in d_{TV} even when exploitability is small, so near-equilibrium play does not pin down a unique marginal. In other words, regret-matching is selecting among multiple equilibria. The learning dynamics determines which equilibrium mixture is reached, and these can induce different one-coordinate marginals. This is consistent with a non-singleton Nash set: different seeds can converge to different equilibrium mixtures, some of which are closer to Hart's target than others.

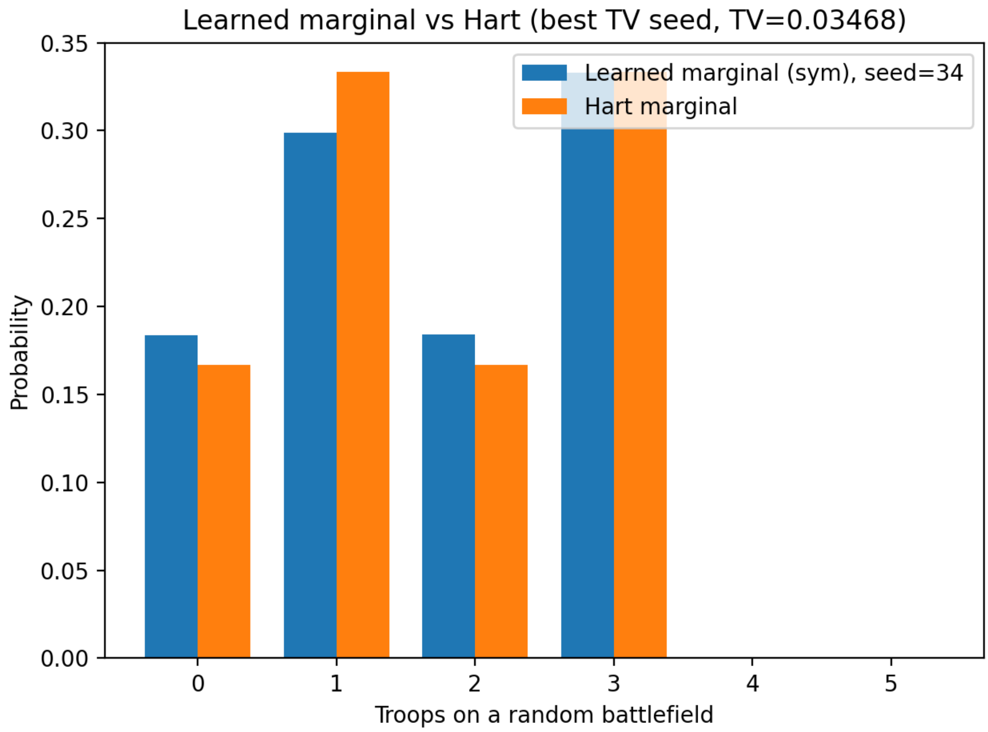

Figure 8 Seed 34: learned marginal versus Hart's target marginal.

Best-TV seed (seed 34). To investigate the smallest TV distance run, we plot the induced marginal for the best seed (seed 34). Compared with Figure 5, the bars align more closely with Hart's target across the main support $\{0, 1, 2, 3\}$, which is consistent with the much smaller TV distance and indicates that this trajectory converged closer to Hart's predicted marginal at this horizon.

Does it match Hart? Overall, regret-matching produces strategies with small exploitability (hence near the Nash set). The induced symmetrised marginals are broadly consistent with Hart's framework, but they do not concentrate around Hart's specific predicted marginal at $T = 10{,}000$. The TV distances remain bounded away from 0 for most seeds. This suggests that Hart's marginal should be interpreted as one equilibrium-consistent marginal among several possible learning outcomes in this instance.

Remark. Since the game is symmetric (same action sets and budgets), Hart's target marginal is a symmetric-equilibrium prediction and therefore applies to either player. We therefore report the comparison for Player 1, an analogous comparison for Player 2 can be performed if desired.

12 Extension: visualising the equilibrium set

Overview. Out of curiosity, we attempt to visualise the geometry of the approximate equilibrium set produced by learning. Each mixed strategy lives in a high-dimensional simplex, so the raw objects are difficult to inspect directly. We therefore apply UMAP to embed the learned strategies into 3D, and colour points by exploitability and by total-variation (TV) distance to Hart's predicted marginal.

12.1 UMAP embeddings

UMAP (Uniform Manifold Approximation and Projection) is a non-linear dimension reduction method that constructs a weighted k-nearest-neighbour graph in the original space and then optimises a low-dimensional embedding to preserve these neighbourhood relations. In our setting, each data point is a mixed strategy $(\bar{p}_T, \bar{q}_T, \sigma(\bar{p}_T), \sigma(\bar{q}_T))$.

Remark. UMAP is primarily a **neighbourhood-preserving** visualisation. The absolute coordinates and global distances in the plot should not be interpreted literally.

Why we embed $\sqrt{p}$ rather than p. Mixed strategies lie on a simplex, so Euclidean distances between raw probability vectors p can be overly driven by a few large entries. We therefore apply the elementwise transform $p \mapsto \sqrt{p}$ before running UMAP. In this $\sqrt{p}$-space, Euclidean distances correspond to the Hellinger geometry on distributions, giving a more meaningful notion of similarity between strategies.

12.2 Unsymmetrised strategies

Figures 9 show UMAP embeddings of the unsymmetrised time-averaged strategies, coloured by exploitability. Two features stand out:

- **No single concentration:** low-exploitability points are not confined to one location; learning yields a spread of near-equilibrium strategies, consistent with a non-singleton Nash set.
- **Diffuse geometry:** the embedding remains broadly scattered, indicating substantial variability across seeds even when exploitability is small.

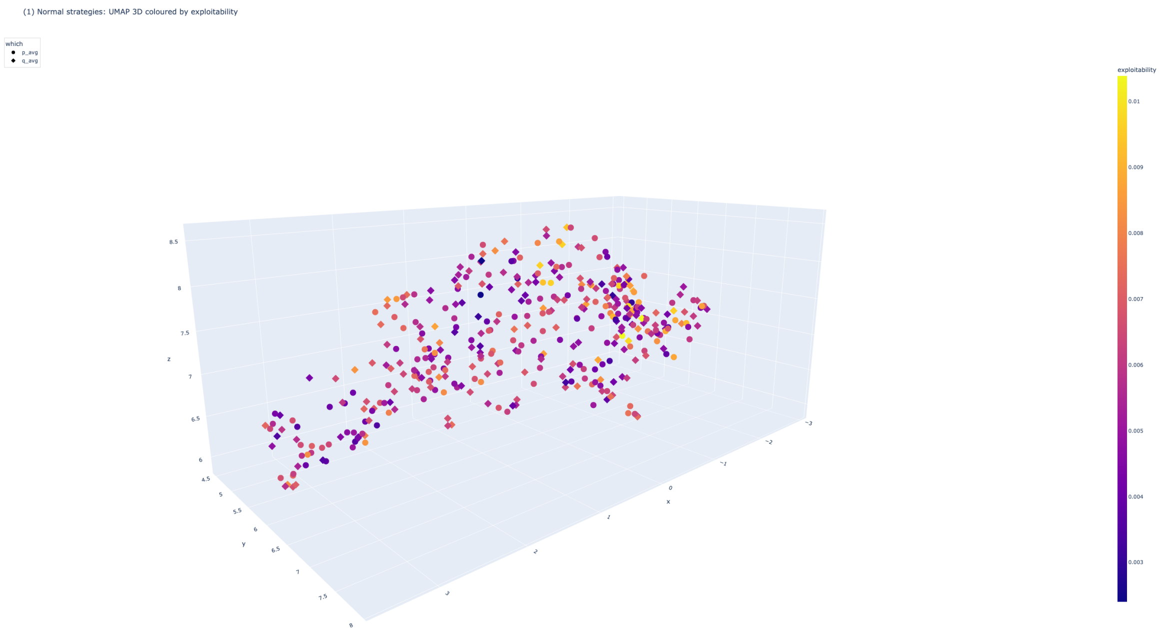

Figure 9 UMAP 3D embedding of unsymmetrised strategies, coloured by exploitability.

12.3 Symmetrised strategies

Figures-10 repeat the visualisation after applying symmetrisation. Compared to the normal strategies, one visual change is apparent.

- **More structured geometry.** Rather than a diffuse cloud, the points lie close to two curved "branches" in the embedding. This is consistent with the idea that enforcing permutation-invariance restricts the effective degrees of freedom, so learning explores a smaller subset of the Nash set.

My hypothesis is that the multiple branches might be an equilibrium set with two families of mixtures that are all equilibrium-consistent under symmetrisation, and different random seeds settle into different regions of this set.

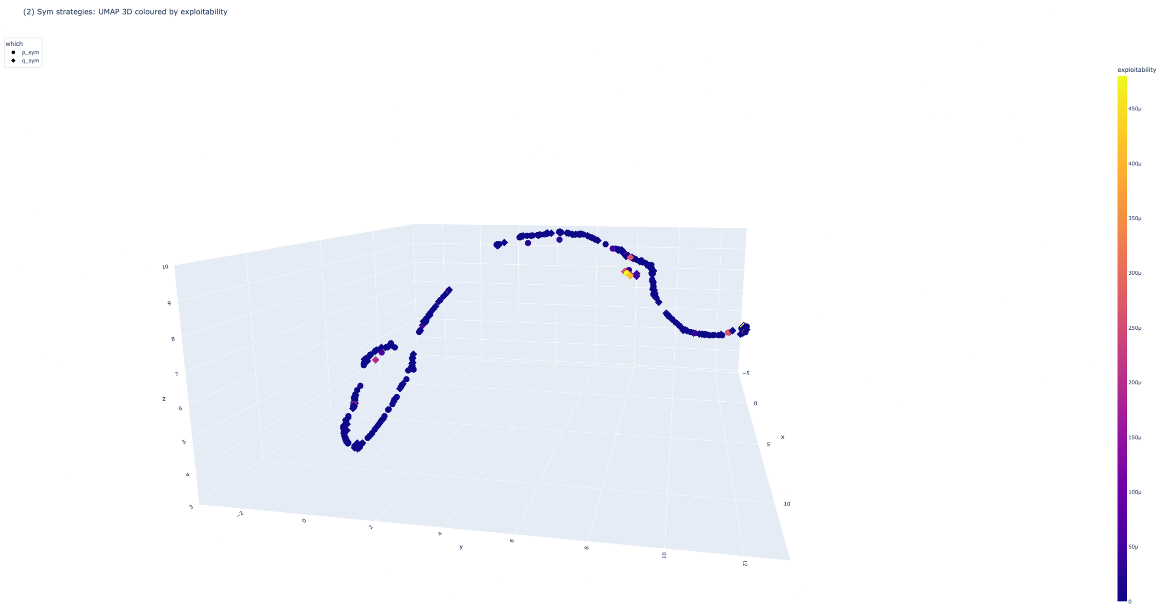

Figure 10 UMAP 3D embedding of symmetrised strategies, coloured by exploitability.

12.4 Symmetrised strategies with Hart anchor (TV distance)

Finally, Figures 11 colour the symmetrised strategies by TV distance to Hart's predicted marginal, and include Hart's marginal as an explicit anchor point (the blotto mixture we used in Subsection 6.1). The embedding shows that the smallest-TV strategies cluster in a local region around the Hart anchor, while other symmetrised near-equilibria lie further away.

This supports the conclusion in Subsection 11.3: Hart's marginal is *equilibrium-consistent* and is approached by a subset of learning runs, but it is not the unique outcome of regret learning in this instance.

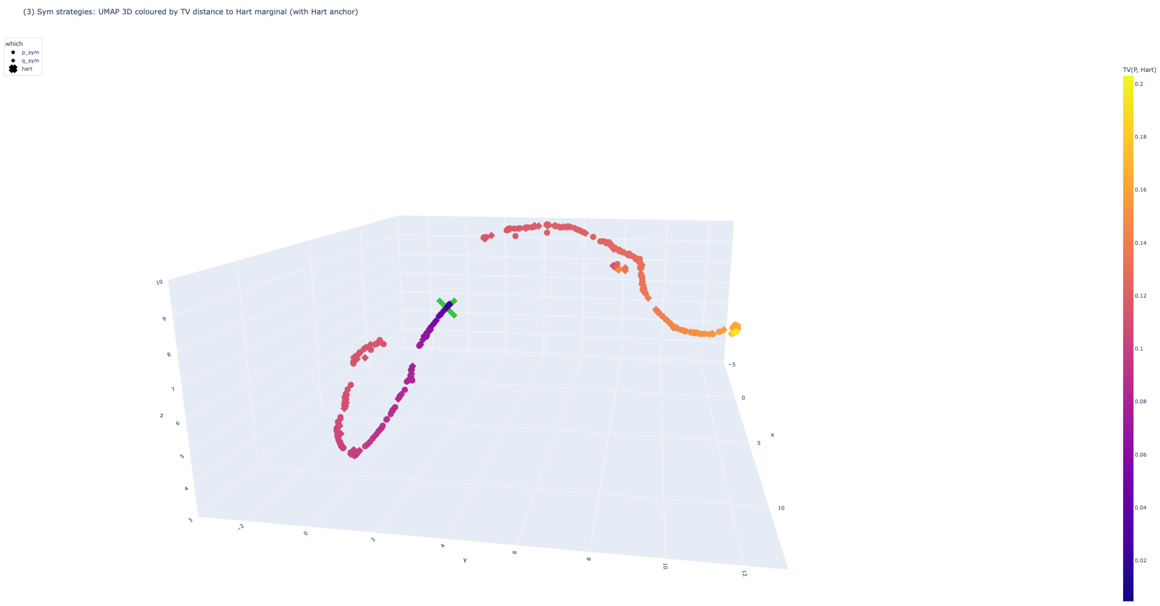

Figure 11 UMAP 3D embedding of symmetrised strategies, coloured by TV distance to Hart's marginal (Hart anchor shown).

Interactive plots. For a more detailed inspection (2D plots and rotating the 3D embeddings), see the accompanying Jupyter notebook in the repository [7].

13 Conclusion

Summary. This project combined theoretical and computational perspectives on the finite Colonel Blotto game.

On the theory side, we reviewed Hart's approach via symmetrisation from [1]: solve the General Lotto relaxation and transfer the solution back to discrete Blotto using feasibility, yielding an equilibrium-consistent *random-battlefield marginal* for $B(5, 5; 3)$.

On the computational side, we implemented sampled regret-matching from Neller–Lanctot [2] for the $(A, K) = (5, 3)$ instance and evaluated convergence using exploitability, as well as total variation distance to compare the learned symmetrised marginal with Hart's target marginal.

Overall, we were able to answer the project questions. Regret-matching's time-averaged strategies approach the Nash set: the estimated game value is close to 0, and exploitability decays at the standard $O(T^{-1/2})$ rate. Applying symmetrisation at evaluation time yields faster empirical decay in this small instance (approximately $O(T^{-1})$), though we do not claim this holds in general. Comparing induced marginals to Hart's target shows that some trajectories converge close to Hart's prediction, but across seeds the outcome is better described as convergence to a *symmetric Nash set* rather than a unique point. The UMAP visualisations further support this non-uniqueness: different runs settle into different low-exploitability strategies. Accordingly, Hart's marginal should be viewed as one equilibrium-consistent marginal within the symmetric set, not the unique outcome of learning.

Next steps. A natural extension is to scale the experiments to larger A and K to test whether the qualitative Hart comparison improves with a finer discrete approximation, and to examine whether different learning dynamics induce different equilibrium selections.

On the theory side, it would be interesting to connect these experiments to results on Blotto equilibria with *continuous* strategy spaces or "infinite" strategy sets in Contests with limited resources in Waging Simple Wars, A Complete Characterization of Two Battlefield Blotto Equilibria [8] and on continuous-allocation variants with limited resources in Contests with limited resources [9], to understand how equilibrium concepts and solution structure change when allocations are not integers.

Algorithmically, we could look into counterfactual regret minimisation CFR-style updates or other regret minimisers with variance reduction, and compare convergence and equilibrium selection under the same metrics.

A Theorems, propositions, remarks

A.1 Discrete General Lotto Solution Theorems taken directly from [1].

Theorem A.1 (Theorem 2). *Let $a \geq b > 0$ where a is an integer. Then the value of the General Lotto game $\Gamma(a,b)$ is*

$$\operatorname{val}\,\Gamma(a,b) \;=\; \frac{a-b}{a} \;=\; 1-\frac{b}{a}.$$

The optimal strategies are as follows:

(i) *When $a=b$ the strategy X is optimal (for either player) if and only if*

$$X \in \operatorname{conv}\{U_O^a,\, U_E^a\}.$$

(ii) *When $a>b$ the strategy U_O^a is the unique optimal strategy of Player 1.*

(iii) *When $a>b$ the strategies*

$$\left(1-\frac{b}{a}\right)\mathbf{1}_0 \;+\; \frac{b}{a}V \qquad \text{with } V \in \operatorname{conv}\{U_O^a,\, U_E^a\}$$

are optimal strategies of Player 2.

(iv) *Every optimal strategy Y of Player 2 satisfies $Y \leq 2a$ and*

$$1-\frac{b}{a} \;\leq\; \mathbb{P}(Y=0) \;\leq\; 1-\frac{b}{a+1}.$$

Theorem A.2 (Theorem 3). *Let $a=m+\alpha$ and $b=m+\beta$ where $m \geq 0$ is an integer and $0<\alpha,\beta<1$. Then the value of the General Lotto game $\Gamma(a,b)$ is*

$$\operatorname{val}\,\Gamma(a,b) \;=\; \frac{a-b}{\lceil a \rceil} \;=\; \frac{\alpha-\beta}{m+1},$$

and the unique optimal strategies are

$$X^\star = (1-\alpha)U_E^m + \alpha U_O^{m+1} \quad \textit{for Player 1} \qquad \textit{and} \qquad Y^\star = (1-\beta)U_E^m + \beta U_O^{m+1} \quad \textit{for Player 2}.$$

Theorem A.3 (Theorem 4). *Let $a = m + \alpha$ and $b \le m$ where $m \ge 1$ is an integer and $0 < \alpha < 1$. Then the value of the General Lotto game $\Gamma(a, b)$ is*

$$\operatorname{val}\Gamma(a,b) = (1-\alpha)\frac{\lfloor a\rfloor - b}{\lfloor a\rfloor} + \alpha\frac{\lceil a\rceil - b}{\lceil a\rceil} = 1 - \frac{(1-\alpha)b}{m} - \frac{\alpha b}{m+1}.$$

The optimal strategies are as follows:

(i) *The strategy*

$$Y^\star = \Big(1 - \frac{b}{m}\Big)\mathbf{1}_0 + \frac{b}{m}U_E^m$$

is the unique optimal strategy of Player 2.

(ii) *The strategy*

$$X^\star = (1-\alpha)U_O^m + \alpha U_O^{m+1}$$

is an optimal strategy of Player 1, and, when $b = m$, so are

$$(1-\alpha)V + \alpha U_O^{m+1} \qquad \textit{for all } V \in \operatorname{conv}\{U_O^m, U_E^m\}.$$

(iii) *Every optimal strategy X of Player 1 satisfies $X \le 2m+1$; moreover, it also satisfies $X \ge 1$ when $b < m$, and*

$$\mathbb{P}(X = 0) \ \le\ \frac{1-\alpha}{m+1} \qquad \textit{when } b = m.$$

A.2 Convergence rate of regret learning taken directly from [3]

Theorem A.4 (Theorem 7.2). *If the row player follows a Hannan-consistent (no–external-regret) strategy, then, regardless of the column player's behaviour,*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) \ \le\ V \qquad \textit{almost surely},$$

where $V = \max_q \min_p \bar{\ell}(p, q)$ is the value of the zero-sum game.

Corollary A.4.1 (Corollary 7.1). *If* both *players follow Hannan-consistent strategies, then the average loss converges to the value:*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(I_t, J_t) \ =\ V \qquad \textit{almost surely}.$$

Remark (Remark 7.4). If both players follow some Hannan consistent strategy, then it is also easy to see that the product distribution $\hat{p}_n \times \hat{q}_n$ formed by the (marginal) empirical distributions of play

$$\hat{p}_{i,n} \ =\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{I_t = i\}} \qquad \text{and} \qquad \hat{q}_{j,n} \ =\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{J_t = j\}}$$

of the two players converges, almost surely, to the set of Nash equilibria $\pi = p \times q$ of the game. However, it is important to note that this does not mean that the players' joint play is close to a Nash equilibrium in the long run. Indeed, one cannot conclude that the *joint empirical frequencies of play*

$$\hat{P}_n(i,j) \ =\ \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{I_t = i,\, J_t = j\}}$$

converge to the set of Nash equilibria. All one can say is that $\hat{P}_n$ converges to the Hannan set of the game (defined later), which, even for zero-sum games, may include joint distributions that are not Nash equilibria.

B Examples

B.1 $B(5,5;3)$ subgame explicitly in matrix form

A 3×3 subgame. Choose three pure allocations in $\mathcal{X}$ (for $(A,K)=(5,3)$):

$$a_1 = (0,0,5), \qquad a_2 = (0,2,3), \qquad a_3 = (1,1,3).$$

With payoff $u(x,y) = \frac{1}{3}\sum_{k=1}^{3} \operatorname{sgn}(x_k - y_k)$, the restricted payoff matrix $M^{(3)} \in \mathbb{R}^{3\times 3}$ with $M^{(3)}_{rs} = u(a_r, a_s)$ is

$$M^{(3)} = \begin{pmatrix} 0 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 \end{pmatrix}.$$

B.2 Computing expected payoff (matrix version) example

Concrete example of expected payoff for Player 1. Let us work with the following pure allocations in $\mathcal{X}$:

$$a_1 = (0,2,3), \qquad a_2 = (1,1,3), \qquad a_3 = (0,0,5),$$

where each vector lists the numbers of troops assigned to the three battlefields.

Recall that $u(x,y)$ denotes the (normalised) Blotto payoff to Player 1 when Player 1 plays the pure allocation x and Player 2 plays the pure allocation y.

From the definition of the payoff matrix M, namely $M_{ij} = u(x^{(i)}, x^{(j)})$, the relevant payoff values are

$$u(a_1,a_2) = 0, \qquad u(a_1,a_3) = 0, \qquad u(a_3,a_2) = -\frac{1}{3}, \qquad u(a_3,a_3) = 0,$$

where, for example,

$$u(a_3,a_2) = \frac{1}{3}\Big(\operatorname{sgn}(0-1) + \operatorname{sgn}(0-1) + \operatorname{sgn}(5-3)\Big) = \frac{1}{3}(-1-1+1) = -\frac{1}{3}.$$

(Interpretation: a_3 loses two battlefields and wins one, so the average outcome is $(-1-1+1)/3$.)

Now consider mixed strategies supported on these actions:

$$x = \frac{1}{2}\,\delta_{a_1} + \frac{1}{2}\,\delta_{a_3}, \qquad y = \frac{1}{4}\,\delta_{a_2} + \frac{3}{4}\,\delta_{a_3}.$$

Player 1 plays a_1 with probability $\frac{1}{2}$ and a_3 with probability $\frac{1}{2}$ and Player 2 plays a_2 with probability $\frac{1}{4}$ and a_3 with probability $\frac{3}{4}$.

Restricting M to rows $\{a_1, a_3\}$ and columns $\{a_2, a_3\}$ gives the 2×2 submatrix

$$M_{\text{sub}} = \begin{pmatrix} u(a_1,a_2) & u(a_1,a_3) \\ u(a_3,a_2) & u(a_3,a_3) \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ -\frac{1}{3} & 0 \end{pmatrix}.$$

In these coordinates we identify the mixed strategies with probability vectors

$$x = (\tfrac{1}{2}, \tfrac{1}{2})^\top, \qquad y = (\tfrac{1}{4}, \tfrac{3}{4})^\top,$$

The expected payoff to Player 1 is

$$x^\top M y \;=\; x^\top M_{\text{sub}} y = -\tfrac{1}{24}.$$

Therefore, when Player 1 plays x and Player 2 plays y, Player 1's expected payoff is

$$x^\top M y \;=\; -\tfrac{1}{24}.$$

(Equivalently, Player 2's expected payoff is $+\frac{1}{24}$ since the game is zero-sum.)

B.3 Inducing a marginal distribution from the symmetrised regret-matching strategy

Example ($A = 5$, $K = 3$). Let $\mathcal{X}$ contain the allocations $(5,0,0)$ and $(3,2,0)$ (among others), and consider the mixed strategy

$$\mu \ = \ 0.6\,\delta_{(5,0,0)} \ + \ 0.4\,\delta_{(3,2,0)}.$$

For $x = (5,0,0)$ the troop count on a uniformly random battlefield equals 5 with probability $1/3$ and 0 with probability $2/3$. For $x = (3,2,0)$ it equals 3, 2, and 0 each with probability $1/3$. Therefore the induced marginal P_μ on $\{0,1,\ldots,5\}$ is

$$P_\mu(0) = 0.6\cdot\frac{2}{3}+0.4\cdot\frac{1}{3} = \frac{8}{15}, \qquad P_\mu(2) = 0.4\cdot\frac{1}{3} = \frac{2}{15}, \qquad P_\mu(3) = 0.4\cdot\frac{1}{3} = \frac{2}{15}, \qquad P_\mu(5) = 0.6\cdot\frac{1}{3} = \frac{1}{5},$$

and $P_\mu(t) = 0$ for $t \in \{1,4\}$. We also have $\sum_{t=0}^{5} P_\mu(t) = 1$.

References

[1] S. Hart. Discrete colonel blotto and general lotto games. *International Journal of Game Theory*, 36(3–4):441–460, 2008.

[2] T. W. Neller and M. Lanctot. An introduction to counterfactual regret minimization. In *Proceedings of the Fourth Symposium on Educational Advances in Artificial Intelligence (EAAI-2013): Model AI Assignments*, volume 11, 2013.

[3] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[4] S. van Strien. Dynamics of learning and iterated games: Lecture notes. Lecture notes, Imperial College London, 2025. MATH60007/70007/97069, 3 October 2025.

[5] G. Farina, J. Grand-Clément, C. Kroer, C.-W. Lee, and H. Luo. Regret matching+:(in) stability and fast convergence in games. *Advances in Neural Information Processing Systems*, 36: 61546–61572, 2023.

[6] F. Timbers, N. Bard, E. Lockhart, M. Lanctot, M. Schmid, N. Burch, J. Schrittwieser, T. Hubert, and M. Bowling. Approximate exploitability: Learning a best response in large games. *arXiv preprint*, 2020.

[7] S. Yap. Dynamic-of-learning-and-iterated-games-project-3, 2025.

[8] S. T. Macdonell and N. Mastronardi. Waging simple wars: a complete characterization of two-battlefield blotto equilibria. *Economic Theory*, 58(1):183–216, 2015.

[9] D. Kvasov. Contests with limited resources. *Journal of Economic Theory*, 136(1):738–748, 2007.