Autonomous Agents and Multi-Agent Systems (2020) 34:7
https://doi.org/10.1007/s10458-019-09432-y

Bounds and dynamics for empirical game theoretic analysis

Karl Tuyls[1] · Julien Perolat[1] · Marc Lanctot[3] · Edward Hughes[2] · Richard Everett[2] · Joel Z. Leibo[2] · Csaba Szepesvári[3] · Thore Graepel[2]

Abstract
This paper provides several theoretical results for empirical game theory. Specifically, we introduce bounds for empirical game theoretical analysis of complex multi-agent interactions. In doing so we provide insights in the empirical meta game showing that a Nash equilibrium of the estimated meta-game is an approximate Nash equilibrium of the true underlying meta-game. We investigate and show how many data samples are required to obtain a close enough approximation of the underlying game. Additionally, we extend the evolutionary dynamics analysis of meta-games using heuristic payoff tables (HPTs) to asymmetric games. The state-of-the-art has only considered evolutionary dynamics of symmetric HPTs in which agents have access to the same strategy sets and the payoff structure is symmetric, implying that agents are interchangeable. Finally, we carry out an empirical illustration of the generalised method in several domains, illustrating the theory and evolutionary dynamics of several versions of the *AlphaGo* algorithm (symmetric), the dynamics of the Colonel Blotto game played by human players on Facebook (symmetric), the dynamics of several teams of players in the capture the flag game (symmetric), and an example of a meta-game in Leduc Poker (asymmetric), generated by the policy-space response oracle multi-agent learning algorithm.

Keywords Empirical games · Asymmetric games · Replicator dynamics

1 Introduction

Using game theory to examine multi-agent interactions in complex systems is a non-trivial task, especially when a payoff table or normal form representation is not directly available. Works by Walsh et al. [39,40], Wellman et al. [43,44], and Phelps et al. [23], have shown the great potential of using heuristic strategies and empirical game theory to examine such

✉ Karl Tuyls
karltuyls@google.com

✉ Julien Perolat
perolat@google.com

[1] DeepMind, Paris, France

[2] DeepMind, London, UK

[3] DeepMind, Edmonton, Canada

Published online: 04 December 2019

interactions at a higher strategic meta-level, instead of trying to capture the decision-making processes at the level of the atomic actions involved. Doing this turns the interaction in a smaller normal form game, or heuristic or meta-game, with the higher-level strategies now being the primitive actions of the game, making the complex multi-agent interaction amenable to game theoretic analysis.

Others have built on this empirical game theoretic methodology and applied these ideas to no limit Texas hold'em Poker and various types of double auctions for example, see [16,22–24,30], showing that a game theoretic analysis at the level of meta-strategies yields novel insights into the type and form of interactions in complex systems.

Major limitations of this empirical game theoretic approach are that it comes without theoretical guarantees on the approximation of the true underlying meta-game (a model of the actual game or interaction) by an estimated meta-game based on sampled data or simulations, and that it is unclear how many data samples are required to achieve a good approximation. Additionally, when examining the evolutionary dynamics of these games the method remains limited to symmetric situations, in which the agents or players have access to the same set of strategies, and are interchangeable. One approach is to ignore asymmetry (types of players), and average over many samples of types resulting in a single expected payoff to each player in each entry of the meta-game payoff table. Many real-world situations though are asymmetric in nature and involve various roles for the agents that participate in the interactions. For instance, buyers and sellers in auctions, or games such as Scotland Yard [21], but also different roles in e.g. robotic soccer (defender vs striker) [29] and even natural language (hearer vs speaker). This type of analysis comes without strong guarantees on the approximation of the true underlying meta-game by an estimated meta-game based on sampled data, and remains unclear about how many data samples are required to achieve a good approximation.

In this paper we address these problems. We use the fact that a Nash equilibrium of the estimated game is a 2ϵ-Nash equilibrium of the underlying meta-game, showing that we can closely approximate the real Nash equilibrium as long as we have enough data samples from which to build the meta-game payoff table. Furthermore, we also examine how many data samples are required to confidently approximate the underlying meta-game. We also show how to generalise the heuristic payoff or meta-game method introduced by Walsh *et al.* to two-population asymmetric games.

Finally, we illustrate the generalised method in several domains. We carry out an experimental illustration on the *AlphaGo* algorithm [27], Colonel Blotto [17], Capture the Flag (CTF) and an asymmetric Leduc poker game. In the *AlphaGo* experiments we show how a symmetric meta-game analysis can provide insights into the evolutionary dynamics and strengths of various versions of the *AlphaGo* algorithm while it was being developed, and how intransitive behaviour can occur by introducing a non-related strategy. In the Colonel Blotto game we illustrate how the methodology can provide insights into how humans play this game, constructing several symmetric meta-games from data collected on Facebook. In the CTF game we examine the dynamics of teams of two agents playing the Capture the Flag game, show examples of intransitive behaviours occurring between these advanced agents and illustrate how Elo rating ([8]) is incapable of capturing such intransitive behaviours. Finally, we illustrate the method in Leduc poker, by examining an asymmetric meta-game, generated by a recently introduced multiagent reinforcement learning algorithm, policy-space response oracles (PSRO) [18]. For this analysis we rely on some theoretical results that connect an asymmetric normal form game to its symmetric counterparts [32].

2 Related work

The purpose of the first applications of empirical game-theoretic analysis (EGTA) was to reduce the complexity of large economic problems in electronic commerce, such as continuous double auctions, supply chain management, market games, and automated trading [39,44]. While these complex economic problems continue to be a primary application area of these methods [5,37,38,41], the general technique has been applied in many different settings. These include analysis interaction among heuristic meta-strategies in poker [24], network protocol compliance [43], collision avoidance in robotics [11], and security games [20,25,48]. Research that followed on Walsh's [39] initial work branched off in two directions: the first strand of work focused on strategic reasoning for simulation-based games [44], while the second strand focused on the evolutionary dynamical analysis of agent behavior inspired by evolutionary game theory [31,33]. The initial paper of Walsh et al. contained innovative ideas that resulted in both research strands taking off in slightly different directions. The current paper is situated in the second line of work focusing on the evolutionary dynamics of empirical or meta-games.

Evolutionary dynamics (foremost replicator dynamics) have often been presented as a practical tool for analyzing interactions among meta-strategies found in EGTA [2,11,39], and for studying the change in policies of multiple learning agents [3], as the EGTA approach is largely based on the same assumptions as evolutionary game-theory, viz. repeated interactions among sub-groups sampled independently at random from an arbitrarily-large population of agents. Also several approaches have investigated the use game-theoretic models, in combination with multi-agent learning, for understanding human learning in multi-agent systems, see e.g. [9,26]. There have also been several uses of EGTA in the context of multiagent reinforcement learning. For example, reinforcement learning can be used to find a best response using an succinct policy representation [15], which can be used to validate equilibria found in EGTA [47], as a regularization mechanism to learn more general meta-strategies than independent learners [18], or to determine the stability of non-adaptive trading strategies such as zero intelligence [49].

A major component of the EGTA paradigm is the estimation of the meta-game that acts as an approximation of the more complex underlying meta-game (like sequential games for example). The quality of the analyses and strategies derived from these estimates depend crucially on the quality of the approximation. The first preferential sampling scheme suggested using an information-theoretic *value of information* criterion to focus the Monte Carlo samples [40]. Other initial approaches to efficient estimation, mentioned in [44], used regression to generalize the payoff of several different complex strategy profiles [36]. Stochastic search methods, such as simulated annealing, were also proposed as means to obtain Nash equilibrium approximations from simulation-based games [35]. More recent work also suggests player reductions that preserve deviations with granular subsampling of the strategy space to get higher-quality information from a finite number of samples [46]. Finally, there is an online tool that helps with managing EGTA experiments [6], which employs a sampling procedure that prioritizes by the estimated regret of the corresponding strategies, which is known to approach the true regret of the underlying game [34]. Despite this, the authors of [6] claim, to the best of their knowledge, that "the construction of optimal sequential sampling procedures for EGTA remains an open question". This work addresses this question of sampling given current estimates and their errors.

3 Preliminaries

In this section, we introduce the necessary background to describe our game theoretic meta-game analysis of the repeated interaction between p players. For the sake of completeness we also provide some theoretical properties of heuristic payoff tables in the appendix of the paper (see “Appendix A”), that have not been treated in the literature before, but point out that this can be easily skipped as the main results can be understood without this section.

3.1 Normal form games

In a p-player Normal Form Game (NFG), players are involved in a single round strategic interaction. Each player $i \in [p] \doteq \{1, \dots, p\}$ chooses a ‘strategy’ $\pi^i \in [k_i]$ from a set of k_i strategies and receives a payoff $r^i(\pi^1, \dots, \pi^p) \in \mathbb{R}$. For the sake of simplicity, we will write $\boldsymbol{\pi}$ for the joint strategy $(\pi^1, \dots, \pi^p) \in [k]^p$ and $\boldsymbol{r}(\boldsymbol{\pi})$ for the joint reward $(r^1(\boldsymbol{\pi}), \dots, r^p(\boldsymbol{\pi}))$. Then a p-player NFG is a tuple $G = (r^1, \dots, r^p)$. Players are also allowed to randomize in which case player i chooses a probability distribution $x^i \in \Delta_{k_i-1} \doteq \{x \in [0,1]^{k_i} : \sum_{j=1}^{k_i} x_j = 1\}$ over $[k_i]$ and the players receive the expected payoff under the joint strategy $\boldsymbol{x} = (x^1, \dots, x^p)$. In particular, player i’s expected payoff is

$$E_{\boldsymbol{\pi}\sim\boldsymbol{x}}[r^i(\pi^1, \dots, \pi^p)] \doteq \sum_{i_1=1}^{k_1} \cdots \sum_{i_p=1}^{k_p} x^1_{i_1} \dots x^p_{i_p} r^i(i_1, \dots, i_p)\,.$$

A symmetric NFG captures interactions where payoffs depend on what strategies are played but not on who plays them. The first condition is therefore that the strategy sets are the same for all players, (*i.e.* $\forall i, j\ k_i = k_j$ and will be written k). The second condition is that if a permutation is applied to the joint strategy $\boldsymbol{\pi}$, the joint payoff is permuted accordingly. Formally, a game G is symmetric if for any permutation σ of $[p]$, we have $\boldsymbol{r}(\boldsymbol{\pi}_\sigma) = \boldsymbol{r}_\sigma(\boldsymbol{\pi})$, where $\boldsymbol{\pi}_\sigma = (\pi^{\sigma(1)}, \dots, \pi^{\sigma(p)})$ and $\boldsymbol{r}_\sigma(\boldsymbol{\pi}) = (r^{\sigma(1)}(\boldsymbol{\pi}), \dots, r^{\sigma(p)}(\boldsymbol{\pi}))$. To repeat, for a game to be symmetric there are two conditions, the players need to have access to the same strategy set and the payoff structure needs to be symmetric, such that players are interchangeable. If one of these two conditions is violated the game is asymmetric.

In the asymmetric case our analysis will focus on the two-player case (two roles) and thus we introduce specific notations for the sake of simplicity. In a two-player normal-form game, each player’s payoff can be seen as a $k_1 \times k_2$ matrix. We will write $A = (a_{uv})_{u\in[k_1], v\in[k_2]}$ for the payoff matrix of player one (*i.e.* $a_{uv} = r^1(u, v)$) and $B = (b_{uv})_{u\in[k_1], v\in[k_2]}$ for the payoff matrix of player two (*i.e.* $b_{uv} = r^2(u, v)$).

In the end, a two player NFG is defined by the tuple $G = (A, B)$.

3.2 Nash equilibrium

In a two-player game, a pair of strategies $(x, y) \in \Delta_{k_1-1} \times \Delta_{k_2-1}$ is a Nash equilibrium of the game (A, B) if no player has an incentive to switch from their current strategy. In other words, (x, y) is a Nash equilibrium if $x^\top Ay = \max Ay$ and $x^\top By = \max x^\top B$, where for a vector u (row-, or column-vector), we define $\max u = \max_i u_i$.

Evolutionary game theory often considers a single strategy x that plays against itself. In this situation, the game is said to have a single population. These situations are often called in the literature single population games. In a single population game, x is a Nash equilibrium if $x^\top Ax = \max Ax$.

3.3 Replicator dynamics

Replicator Dynamics are one of the central concepts from Evolutionary Game Theory [10,12,19,42,50,51]. They describe how a population of replicators, or a strategy profile, evolves in the midst of others through time under evolutionary pressure. Each replicator in the population is of a certain type, and they are randomly paired in interaction. Their reproductive success is determined by their fitness, which results from these interactions. The replicator dynamics express that the population share of a certain type will increase if the replicators of this type have a higher fitness than the population average; otherwise their population share will decrease. This evolutionary process is described according to a first order dynamical system. In a two-player NFG (A, B), the replicator equations are defined as follows:

$$\dot{x}_u = x_u \left((Ay)_u - x^\top Ay \right), \qquad \dot{y}_v = y_v \left((x^\top B)_v - x^\top By \right) \tag{1}$$

with $\mathbf{x} \in \Delta_{k_1-1}, \mathbf{y} \in \Delta_{k_2-1}$. The dynamics defined by these two coupled differential equations changes the strategy profile to increase the probability of the strategies that have the best return or are the *fittest*.

In the case of a symmetric two-player game ($A = B^\top$), the replicator equations assume that both players play the same strategy profile (*i.e.* player one and two play according to x) and the dynamics are defined as follows:

$$\dot{x}_l = x_l \left((Ax)_l - x^\top Ax \right) \tag{2}$$

3.4 Meta games and heuristic payoff tables

A meta game (or empirical game) is a simplified model of a complex multi-agent interaction. In order to analyze complex multi-agent systems like poker, we do not consider all possible atomic actions but rather a set of relevant meta-strategies that are often played [24]. These meta strategies (or sometimes styles of play), over atomic actions, are commonly played by players such as for instance "passive/aggressive" or "tight/loose" in poker. A p-type meta game is now a p-player repeated NFG where players play a limited number of meta strategies. Following our poker example, the strategy set of the meta game will now be defined as the set {"*aggressive*", "*tight*", "*passive*"} and the reward function as the outcome of a game between p-players using different profiles.

When a NFG representation of such a complex multi-agent interaction is not available, one can use the heuristic payoff table (HPT), as introduced in Walsh et al. [39,40]. The idea of the HPT is to capture the expected payoff of high-level meta-strategies through simulation, or from data of interactions, when the payoffs are not readily available (e.g. through a given NFG). Note that the purpose of the HPT is not to directly apply it to simple known matrix games - in that case one can just plug the normal form game directly in the replicator equations. Continuous-time replicator dynamics assume an infinite population, which is approximated in the HPT method by a finite population of p individuals to be able to run simulations. As such, the HPT is only an approximation. The larger p gets, the more subtleties are captured by the HPT and the resulting dynamics will be more accurately reflecting the underlying true dynamics.

Table 1 An example of a meta game payoff table

$$M = \left(\begin{array}{ccc|ccc} N_{i1} & N_{i2} & N_{i3} & U_{i1} & U_{i2} & U_{i3} \\ \hline 6 & 0 & 0 & 0 & 0 & 0 \\ & \dots & & & \dots & \\ 4 & 0 & 2 & -0.5 & 0 & 1 \\ & \dots & & & \dots & \\ 0 & 0 & 6 & 0 & 0 & 0 \end{array}\right)$$

If we were to construct a classical payoff table for $\mathbf{r}$ we would require k^p entries in the NFG table and this becomes large very quickly. Since all players can choose from the same strategy set and all players receive the same payoff for being in the same situation, we can simplify our payoff table. This means we consider a game where the payoffs for playing a particular strategy depend only on the other strategies employed by the other players, but not on who is playing them. This corresponds to the setting of symmetric games.

We now introduce the HPT. Let N be a matrix, where each row N_i is a vector of counts $(n_1, \dots, n_k)$ where $\sum_j n_j = p$: n_j indicates how many of the p players play strategy j. The number of such distinct count vectors (which we also view as a discrete distribution) can be shown to be $m = \binom{p+k-1}{p}$, which is the number of rows of N. Each distribution over strategies can be simulated (or derived from data), returning a vector of expected rewards $u(N_i)$ (one for each of the k strategies). Let U be an $m \times k$ matrix which captures the payoffs corresponding to the rows in N, i.e., $U_i = u(N_i)$. We refer to an HPT as $M = (N, U)$. Note that normalizing a count vector $(n_1, \dots, n_k)$ by dividing it by p gives a probability vector $\boldsymbol{x} = (n_1/p, \dots, n_k/p)$, which we call a discrete strategy distribution.

Suppose we have a meta-game with 3 meta-strategies ($k = 3$) and 6 players ($p = 6$) that interact in a 6-type game, this leads to a meta game payoff table of 28 entries (which is a good reduction from 3^6 cells). An important advantage of this type of table is that it easily extends to many agents, as opposed to the classical payoff matrix. Table 1 provides an example for three strategies and three players. The left-hand side shows the counts and gives the matrix N, while the right-hand side gives the payoffs for playing any of the strategies given the discrete profile and corresponds to matrix U.

The HPT has one row per possible discrete distribution, for each row we usually run many simulations (or collect many data samples) to determine the expected payoff of each type present in the discrete distribution. There are p (finite) individuals present in the simulation at all times. In other words we simulate populations of p agents, and record their expected utilities in the HPT.

4 Method

There are now two possibilities, either the meta-game is symmetric, or it is asymmetric. We will start with the simpler symmetric case, which has been studied in empirical game theory, then we continue with asymmetric games, in which we consider two populations, or roles.

4.1 Symmetric meta games

We consider a set of agents or players A with $|A| = n$ that can choose a strategy from a set S with $|S| = k$ and can participate in one or more p-type meta-games with $p \leq n$. If the game is symmetric then the formulation of meta strategies has the advantage that the payoff for a strategy does not depend on which player has chosen that strategy and consequently the payoff for that strategy only depends on the composition of strategies it is facing in the game and not on who is playing the strategy. This symmetry has been the main focus of the use of empirical game theory analysis [22,24,39,44].

In order to analyse the evolutionary dynamics of high-level meta-strategies, we also need to estimate the expected payoff of such strategies relative to each other. In evolutionary game theoretic terms, this is the relative fitness of the various strategies, dependent on the current frequencies of those strategies in the population.

In order to approximate the payoff for an arbitrary mix of strategies in an infinite population of replicators distributed over the species according to $\mathbf{x}$, p individuals are drawn randomly from the distribution $\mathbf{x}$. Let the set of all discrete profiles be denoted by $\omega = \{(p, 0, .., 0), .., (0, .., 0, p)\}$ and let $\mu_i = \{N \in \omega | N_i = 0\}$ where strategy i is not played, and $\bar{\mu_i} = \{N \in \omega | N_i \neq 0\}$ its complement. The probability for selecting a specific row N_i can be computed from $\mathbf{x}$ and N_i as (where $\binom{p}{N_{i1}, N_{i2}, \ldots, N_{ik}}$ is a multinomial coefficient)

$$P(N_i|\mathbf{x}) = \binom{p}{N_{i1}, N_{i2}, \ldots, N_{ik}} \prod_{j=1}^{k} x_j^{N_{ij}}.$$

The expected payoff of strategy π^j, $r^j(\mathbf{x})$, is then computed as the weighted combination of the payoffs given in all rows:

$$r^j(\mathbf{x}) = \frac{\sum_{N_i \in \bar{\mu_j}} P(N_i|\mathbf{x}) U_{ij}}{1 - \sum_{N_i \in \mu_j} P(N_i|\mathbf{x})}.$$

We need to re-normalize (denominator) by ignoring rows that do not contribute to the payoff of a strategy because it is not present in the distribution N_j in the meta-game payoff table. This expected payoff function can now be used in Eq. 2 to compute the evolutionary population change according to the replicator dynamics by replacing $(Ax)_i$ by $r^i(\mathbf{x})$.

If the HPT approach were applied to capture simple matrix games (which is not its purpose), one needs to take into account that in the single population replicator dynamics model, two individuals are randomly matched to play the normal form game. For an infinite population, sampling two individuals with or without replacement is identical. However, for finite populations, and especially if n is small, there is an important difference between sampling with and without replacement. The payoff calculation method shown above correctly reproduces the expected payoff of matrix games if n is sufficiently large for both sampling with replacement and without, however, for smaller n sampling with replacement will result in lower errors. See "Appendix A" for an example and more theoretical properties about HPTs.

4.2 Asymmetric meta games

One can now wonder how the previously introduced method extends to asymmetric games, which has not been considered in the literature. An example of an asymmetric game is the famous battle of the sexes game illustrated in Table 2. In this game both players do have the same strategy sets, i.e., go to the opera or go to the movies, however, the corresponding payoffs for each are different, expressing the differences in preferences that both players have.

If we aim to carry out a similar evolutionary analysis as in the symmetric case, restricting ourselves to two populations or roles, we will need two meta game payoff tables, one for each player over its own strategy set. We will also need to use the asymmetric version of the

Table 2 Battle of the Sexes game: strategies O and M correspond to going to the Opera and going to the Movies respectively

	O	M
O	3, 2	0, 0
M	0, 0	2, 3

Table 3 General 3 × 3 normal form game

	C_1	C_2	C_3
R_1	r_{11}, c_{11}	r_{12}, c_{12}	r_{13}, c_{13}
R_2	r_{21}, c_{21}	r_{22}, c_{22}	r_{23}, c_{23}
R_3	r_{31}, c_{31}	r_{32}, c_{32}	r_{33}, c_{33}

Table 4 An example of an asymmetric meta game payoff table

$$P = \left(\begin{array}{ccc|ccc} N_{i1,j1} & N_{i2,j2} & N_{i3,j3} & U_{i1,j1} & U_{i2,j2} & U_{i3,j3} \\ \hline (1,1) & 0 & 0 & (r_{11}, c_{11}) & 0 & 0 \\ & \cdots & & & \cdots & \\ (1,0) & (0,1) & 0 & (r_{12}, 0) & (0, c_{12}) & 0 \\ (0,1) & (1,0) & 0 & (0, c_{21}) & (r_{21}, 0) & 0 \\ & \cdots & & & \cdots & \\ 0 & 0 & (1,1) & 0 & 0 & (r_{33}, c_{33}) \end{array} \right)$$

Table 5 A decomposed asymmetric meta payoff table for Player 1

$$P = \left(\begin{array}{ccc|ccc} N_{i1,j1} & N_{i2,j2} & N_{i3,j3} & U_{i1,j1} & U_{i2,j2} & U_{i3,j3} \\ \hline 2 & 0 & 0 & r_{11} & 0 & 0 \\ & \cdots & & & \cdots & \\ 1 & 1 & 0 & r_{12} & r_{21} & 0 \\ & \cdots & & & \cdots & \\ 0 & 0 & 2 & 0 & 0 & r_{33} \end{array} \right)$$

replicator dynamics as shown in Eq. 1. Additionally, in order to compute the right payoffs for every situation we will have to interpret a discrete strategy profile in the meta-table slightly different. Suppose we have a 2-type meta game, with three strategies in each player's strategy set. We introduce a generalisation of our meta-table for both players by means of an example shown in Table 4, which corresponds to the general NFG shown in Table 3.

Let's have a look at the first entry in Table 4, i.e., $[(1, 1), 0, 0]$. This entry means that both agents (i and j) are playing their first strategy, expressed by $N_{i1,j1}$, meaning the number of agents N_{i1} playing strategy π_i^1 in the first population equals 1 and that the number of agents N_{j1} playing strategy π_j^2 in the second population equals 1 as well. The corresponding payoff for each player $U_{i1,j1}$ equals (r_{11}, c_{11}). Now lets have a look at the discrete profiles: $[(1, 0), (0, 1), 0]$ and $[(0, 1), (1, 0), 0]$. The first one means that the first player is playing its first strategy while the second player is playing their second strategy. The corresponding payoffs are r_{12} for the first player and c_{12} for the second player. The profile $[(0, 1), (1, 0), 0]$ shows the reverted situation in which the second player plays his first strategy and the first player plays his second strategy, yielding payoffs r_{21} and c_{21} for the first player and second player respectively. In order to turn the table into a similar format as for the symmetric case, we can now introduce p meta-tables, one for each player. More precisely, we get Tables 5 and 6 for players 1 and 2 respectively.

Table 6 A decomposed asymmetric meta payoff table for Player 2

$$Q = \left(\begin{array}{ccc|ccc} N_{i1,j1} & N_{i2,j2} & N_{i3,j3} & U_{i1,j1} & U_{i2,j2} & U_{i3,j3} \\ \hline 2 & 0 & 0 & c_{11} & 0 & 0 \\ & \cdots & & & \cdots & \\ 1 & 1 & 0 & c_{12} & c_{21} & 0 \\ & \cdots & & & \cdots & \\ 0 & 0 & 2 & 0 & 0 & c_{33} \end{array}\right)$$

One needs to take care in correctly interpreting these tables. Let's have a look at row [1, 1, 0] for instance. This should now be interpreted in two ways: one, the first player plays his first strategy while the other player plays his second strategy and he receives a payoff of r_{12}, two, the first player plays his second strategy while the other player plays his first strategy and receives a payoff of r_{21}. The expected payoff $r^i(\mathbf{x})$ can now be estimated in the same way as explained for the symmetric case as we will be relying on symmetric replicator dynamics by decoupling asymmetric games in their *symmetric counterparts* (explained in the next section).

4.3 Linking symmetric and asymmetric games

Here we repeat an important result on the link between an asymmetric game and its symmetric counterpart games. For a full treatment and discussion of these results see [32]. This work proves that if x, y is a Nash equilibrium of the bimatrix game (A, B) (where x and y have the same support[1]), then y is a Nash equilibrium of the single population, or symmetric, game A and x is a Nash equilibrium of the single population, or symmetric, game $B^\top$. Both symmetric games are called the *counterpart games* of the asymmetric game (A, B). The reverse is also true: If y is a Nash equilibrium of the single population game A and x is a Nash equilibrium of the single population game $B^\top$ (and if x and y have the same support), then x, y is a Nash equilibrium of the game (A, B). In our empirical analysis, we use this property to analyze an asymmetric games (A, B) by looking at the counterpart single population games A and $B^\top$. Formally (from [32]):

Theorem 1 *Strategies x and y constitute a Nash equilibrium of an asymmetric game $G = (A, B)$ with the same support (i.e. $I_x = I_y$) if and only if x is a Nash equilibrium of the single population game B^T, y is a Nash equilibrium of the single population game A and $I_x = I_y$. Where $I_x = \{i \mid x_i > 0\}$ and $I_y = \{i \mid y_i > 0\}$.*

The result we state here is limited to strategies with the same support, but this condition can be softened (see [32]).

5 Theoretical insights

As illustrated in the previous section the procedure for empirical meta-game analysis consists of two parts. Firstly, one needs to construct an empirical meta-game utility function for each player. This step can be performed using logs of interactions between players, or by playing the game sufficiently enough (simulations). Secondly, one expects that analyzing the estimated empirical game will give insights in the true underlying empirical game itself (i.e. the game

[1] x and y have the same support if $I_x = I_y$ where $I_x = \{i \mid x_i > 0\}$ and $I_y = \{i \mid y_i > 0\}$.

from which we sample). This section provides insights in the following: how much data is enough to generate a good approximation of the true underlying empirical game? Is uniform sampling over actions or strategies the right method?

5.1 Main lemma

Sometimes players receive a stochastic reward $R^i(\pi^1, \ldots, \pi^p)$ for a given joint action $\boldsymbol{\pi}$. The underlying game we study is $r^i(\pi^1, \ldots, \pi^p) = E\left[R^i(\pi^1, \ldots, \pi^p)\right]$ and for the sake of simplicity the joint action of every player but player i will be written $\boldsymbol{\pi}^{-i}$. In the next two definitions we recall the concept of a Nash equilibria and ϵ-Nash equilibria in p-player games.

Definition A joint strategy $\boldsymbol{x} = (x^1, \ldots, x^p) = (x^i, \boldsymbol{x}^{-i})$ is a Nash equilibrium if for all i:

$$E_{\boldsymbol{\pi}\sim\boldsymbol{x}}\left[r^i(\boldsymbol{\pi})\right] = \max_{\pi^i} E_{\boldsymbol{\pi}^{-i}\sim\boldsymbol{x}^{-i}}\left[r^i(\pi^i, \boldsymbol{\pi}^{-i})\right]$$

Definition A joint strategy $\boldsymbol{x} = (x^1, \ldots, x^p) = (x^i, \boldsymbol{x}^{-i})$ is an ϵ-Nash equilibrium if for all i:

$$\max_{\pi^i} E_{\boldsymbol{\pi}^{-i}\sim\boldsymbol{x}^{-i}}\left[r^i(\pi^i, \boldsymbol{\pi}^{-i})\right] - E_{\boldsymbol{\pi}\sim\boldsymbol{x}}\left[r^i(\boldsymbol{\pi})\right] \leq \epsilon$$

When running an analysis on a meta game, we do not have access to the average reward function $r^i(\pi^1, \ldots, \pi^p)$ but to an empirical estimate $\hat{r}^i(\pi^1, \ldots, \pi^p)$. The following lemma shows that a Nash equilibrium for the empirical game $\hat{r}^i(\pi^1, \ldots, \pi^p)$ is an 2ϵ-Nash equilibrium for the game $r^i(\pi^1, \ldots, \pi^p)$ where $\epsilon = \sup_{\boldsymbol{\pi},i} |\hat{r}^i(\boldsymbol{\pi}) - r^i(\boldsymbol{\pi})|$. A more general statement can be found in [34] (see comment below).

Lemma *If $\boldsymbol{x}$ is a Nash equilibrium for $\hat{r}^i(\pi^1, \ldots, \pi^p)$, then it is a 2ϵ-Nash equilibrium for the game $r^i(\pi^1, \ldots, \pi^p)$ where $\epsilon = \sup_{\boldsymbol{\pi},i} |r^i(\boldsymbol{\pi}) - \hat{r}^i(\boldsymbol{\pi})|$.*

Proof Fix some player index $1 \leq i \leq p$. We extend the notation so that $r^i(\pi^i, \boldsymbol{x}^{-i}) = E_{\boldsymbol{\pi}^{-i}\sim\boldsymbol{x}^{-i}} r^i(\pi^i, \boldsymbol{\pi}^{-i})$ and $r^i(\boldsymbol{x}) = E_{\boldsymbol{\pi}\sim\boldsymbol{x}} r^i(\boldsymbol{\pi})$. We use similar notation with $\hat{r}^i$. Note that it follows from our conditions on r^i and $\hat{r}^i$ that for any π^i, $r^i(\pi^i, \boldsymbol{x}^{-i}) - \hat{r}^i(\pi^i, \boldsymbol{x}^{-i}) \leq \epsilon$ and $\hat{r}^i(\boldsymbol{x}) - r^i(\boldsymbol{x}) \leq \epsilon$. Then,

$$\begin{aligned}
\max_{\pi^i} r^i(\pi^i, \boldsymbol{x}^{-i}) &= \max_{\pi^i} \hat{r}^i(\pi^i, \boldsymbol{x}^{-i}) + \underbrace{\max_{\pi^i} r^i(\pi^i, \boldsymbol{x}^{-i}) - \hat{r}^i(\pi^i, \boldsymbol{x}^{-i})}_{\leq \epsilon} \\
&\leq \hat{r}^i(\boldsymbol{x}) + \epsilon \\
&= r^i(\boldsymbol{x}) + \underbrace{\hat{r}^i(\boldsymbol{x}) - r^i(\boldsymbol{x})}_{\leq \epsilon} + \epsilon \leq r^i(\boldsymbol{x}) + 2\epsilon\,.
\end{aligned}$$

Since i was arbitrary, the result follows. □

This lemma shows that if one can control the difference between $|r^i(\boldsymbol{\pi}) - \hat{r}^i(\boldsymbol{\pi})|$ uniformly over players and actions, then an equilibrium for the empirical game $\hat{r}^i(\pi^1, \ldots, \pi^p)$ is almost an equilibrium for the game defined by the average reward function $r^i(\pi^1, \ldots, \pi^p)$. It is worth mentioning that a related result [34] proves that the set of ϵ-Nash equilibria of the empirical game includes the set of Nash equilibria in the underlying game. Note that in [34] they prove a general result that coincides with ours when $\delta = 0$ and $\epsilon(r) = 0$ in theorem 6.1.

5.2 Finite sample analysis

This section details some concentration results. In practice, we often have access to a batch of observations of the underlying game. We will run our analysis on an empirical estimate of the game denoted by $\hat{r}^i(\boldsymbol{\pi})$. The question then will be either with which confidence can we say that a Nash equilibrium for $\hat{\boldsymbol{r}}$ is a 2ϵ-Nash equilibrium, or for a fixed confidence, for which ϵ can we say that a Nash equilibrium for $\hat{\boldsymbol{r}}$ is a 2ϵ-Nash equilibrium for $\boldsymbol{r}$. In the case we have access to game play, the question is how many samples n do we need to assess that a Nash equilibrium for $\hat{\boldsymbol{r}}$ is a 2ϵ-Nash equilibrium for $\boldsymbol{r}$ for a fixed confidence and a fixed ϵ. For the sake of simplicity, we will assume that the random payoffs are bounded in [0, 1].

5.2.1 The batch scenario

Here we assume that we are given $n(i, \boldsymbol{\pi})$ independent samples to compute the empirical average $\hat{r}^i(\boldsymbol{\pi})$. For $i, \boldsymbol{\pi}$ fixed, Hoeffding's inequality gives that outside of some failure event $\mathcal{E}_{i,\boldsymbol{\pi},\delta}$ whose probability is bounded by δ, $|\hat{r}^i(\boldsymbol{\pi}) - r^i(\boldsymbol{\pi})| \leq \sqrt{\log(2/\delta)/(2n(i,\boldsymbol{\pi}))}$. It follows that outside of the event $\mathcal{E} = \cup_{i,\boldsymbol{\pi}\in S}\mathcal{E}_{i,\boldsymbol{\pi},\delta/(|S|p)}$ with $P(\mathcal{E}) \leq |S|p\frac{\delta}{|S|p} = \delta$,

$$\max_{i,\boldsymbol{\pi}} |\hat{r}^i(\boldsymbol{\pi}) - r^i(\boldsymbol{\pi})| \leq \max_{i,\boldsymbol{\pi}} \sqrt{\frac{\log(2p|S|) + \log(1/\delta)}{2n(i,\boldsymbol{\pi})}}.$$

5.2.2 Uniform sampling

Setting $n(i, \boldsymbol{\pi}) = n/(|S|p)$, the previous bound becomes

$$\max_{i,\boldsymbol{\pi}} |\hat{r}^i(\boldsymbol{\pi}) - r^i(\boldsymbol{\pi})| \leq \sqrt{|S|p\frac{\log(2p|S|) + \log(1/\delta)}{2n}}.$$

It follows, that to guarantee a uniform error of size ϵ with probability $1 - \delta$, it is sufficient if

$$n \geq \frac{(\log(2p|S|) + \log(1/\delta))\, p|S|}{2\epsilon^2}.$$

Our bound is slightly better than the one of [34]. In our case, the probability of the equilibrium of the meta-game to be worse than a 2ϵ-Nash equilibrium is smaller than $2p|S|\exp(-2\epsilon^2 n)$ (where n is the number of samples used) whilst [34] shows an upper-bound of $(K+1)p|S|\exp(-2\epsilon^2 n)$ where K is the maximum number of strategies available per players. In other words, our bound is better than the one from [34] by a multiplicative factor K. That work is most closely related to this paper, but also other finite sample analyses have been proposed in the literature in other contexts [14,45].

6 Experiments

This section presents experiments that illustrate the meta-game approach and its feasibility for examining strengths and weaknesses of higher-level strategies in various domains, including AlphaGo, Colonel Blotto, CTF and the meta-game generated by PSRO. Note that we restrict the meta-games to three strategies here, as we can nicely visualise this in a phase plot, and these still provide useful information about the dynamics in the full strategy spaces.

Table 7 Meta-game payoff table generated from Table 9 in [27] for strategies $\alpha_{rvp}, \alpha_{vp}, \alpha_{rp}$

α_{rvp}	α_{vp}	α_{rp}	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	0.95	0	0.05
0	2	0	0	0.5	0
1	1	0	0.99	0.01	0
0	0	2	0	0	0.5
0	1	1	0	0.39	0.61

The first step of the analysis is always the derivation of the meta-game payoff table itself, for which a sufficiently large data-set is required that allows for computing the relative payoffs of the various strategies under study against each other. We start with the AlphaGo data, continue with colonel Blotto and CTF, and end with examining the asymmetric PSRO game. Also note that these tables are relatively small, as $p = 2$ and $n = 3$, so in these cases we could also use the NFG representation in the replicator dynamics equations. In fact the dynamics of zero-sum 2-player games give equivalent results for NFG and HPT representations (and as such both representations lead to the same visualizations), see the Appendix for more details and proofs. For consistency purposes we therefore stick to HPT representations here.

6.1 AlphaGo

The data set under study consists of 7 *AlphaGo* variations and a number of different Go strategies such as Crazystone and Zen (previously the state-of-the-art). α stands for the algorithm and the indexes r, v, p for the use of respectively *rollouts*, *value nets* and *policy nets* (e.g. α_{rvp} uses all 3). For a detailed description of these strategies see [27]. The meta-game under study here concerns a 2-type NFG with $|S| = 9$. We will look at various 2-faces of the larger simplex. Table 9 in [27] summarises all wins and losses between these various strategies (meeting several times), from which we can compute meta-game payoff tables.

6.1.1 Experiment 1: Strong strategies

This first experiment examines three of the strongest *AlphaGo* strategies in the data-set, i.e., $\alpha_{rvp}, \alpha_{vp}, \alpha_{rp}$. As a first step we created a meta-game payoff table involving these three strategies, by looking at their pairwise interactions in the data set (summarised in Table 9 of [27]). This set contains data for all strategies on how they interacted with the other 8 strategies, listing the win rates that strategies achieved against one another (playing either as white or black) over several games. The meta-game payoff table derived for these three strategies is described in Table 7.

In Fig. 1 we have plotted the directional field of the meta-game payoff table using the replicator dynamics for a number of strategy profiles $\mathbf{x}$ in the simplex strategy space. From each of these points in strategy space an arrow indicates the direction of flow, or change, of the population composition over the three strategies. Figure 2 shows a corresponding trajectory plot. From these plots one can easily observe that strategy α_{rvp} is a strong attractor and consumes the entire strategy space over the three strategies. This restpoint is also a Nash equilibrium. This result is in line with what we would expect from the knowledge we have of the strengths of these various learned policies. Still, the arrows indicate how the

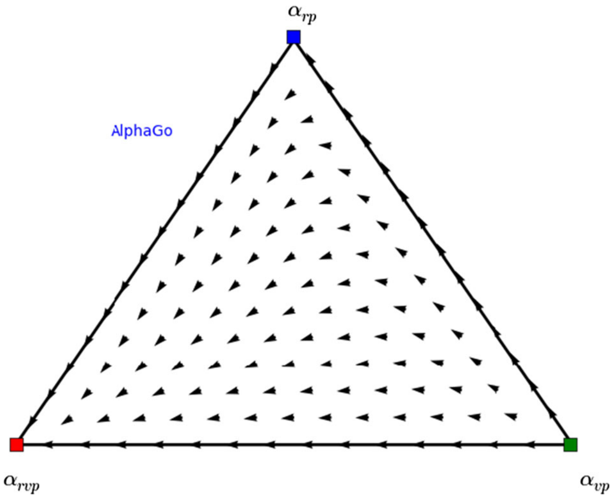


Fig. 1 Directional field plot for the 2-face consisting of strategies $\alpha_{rvp}, \alpha_{vp}, \alpha_{rp}$

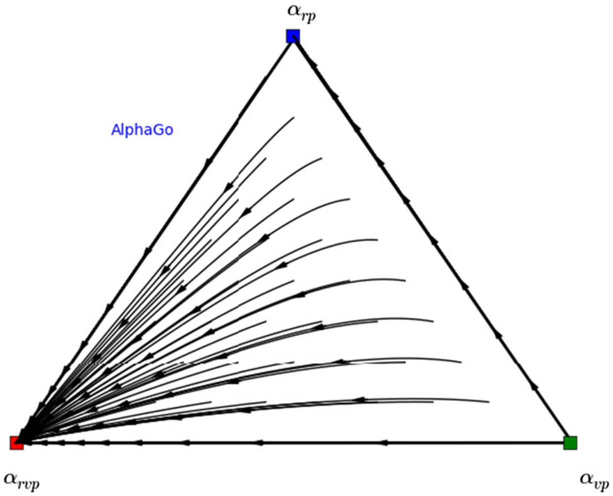


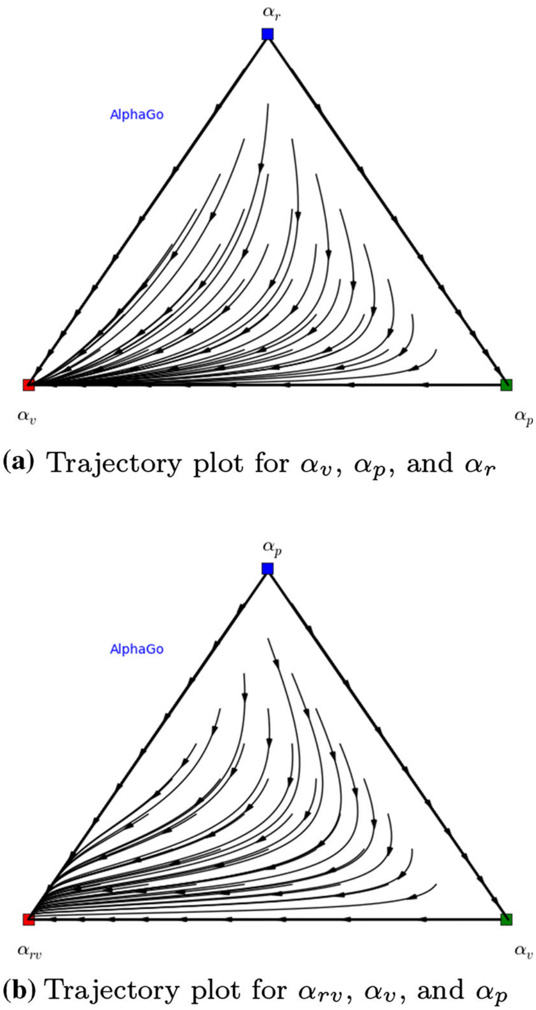
Fig. 2 Trajectory plot for the 2-face consisting of strategies $\alpha_{rvp}, \alpha_{vp}, \alpha_{rp}$

strategy landscape flows into this attractor and therefore provides useful information as we will discuss later.

6.1.2 Experiment 2: Evolution and transitivity of strengths

We start by investigating the 2-face simplex involving strategies α_{rp}, α_{vp} and α_{rv}, for which we created a meta-game payoff table similarly as in the previous experiment (not shown). The evolutionary dynamics of this 2-face can be observed in Fig. 4a. Clearly strategy α_{rp} is a strong attractor and beats on average the two other strategies. We now replace this attractor by strategy α_{rvp} and plot its evolutionary dynamics in Fig. 4b. What can be observed from both trajectory plots in Fig. 4 is that the curvature is less pronounced in plot 4b than it is

Fig. 3 AlphaGo evolutionary dynamics plots for α_v, α_p, α_r, and α_{rv}



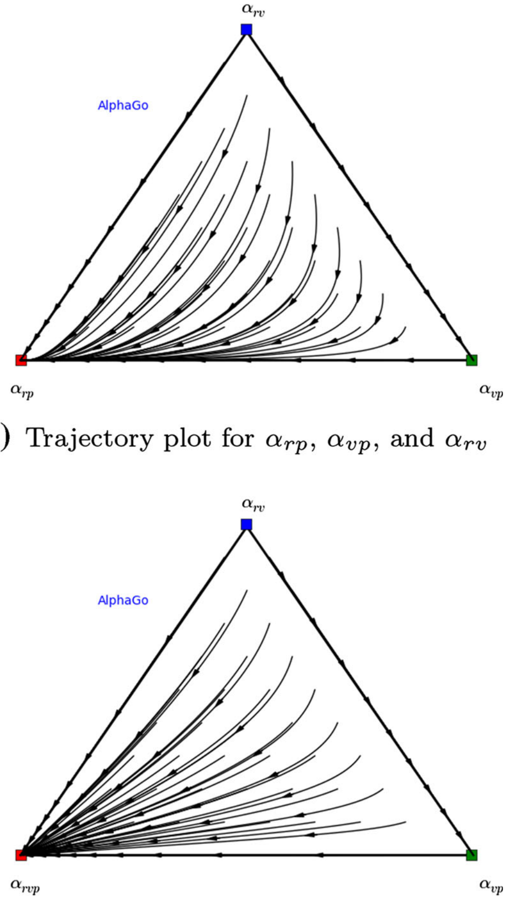
(a) Trajectory plot for α_v, α_p, and α_r

(b) Trajectory plot for α_{rv}, α_v, and α_p

in plot 4a. The reason for this is that the difference in strength between α_{rv} and α_{vp} is less obvious in the presence of an even stronger attractor than α_{rp}. This means that α_{rvp} is now pulling much stronger on both α_{rv} and α_{vp} and consequently the flow goes more directly to α_{rvp}. So even when a strategy space is dominated by one strategy, the curvature (or curl) is a promising measure for the strength of a meta-strategy.

What is worthwhile to observe from the *AlphaGo* dataset, and illustrated as a series in Figs. 3 and 4, is that there is clearly an incremental increase in the strength of the *AlphaGo* algorithm going from version α_r to α_{rvp}, building on previous strengths, without any intransitive behaviour occurring, when only considering a strategy space formed by the *AlphaGo* versions.

Finally, as discussed in Sect. 5, we can now examine how good of an approximation an estimated game is. In the *AlphaGo* domain we only do this analysis for the games displayed in Fig. 4a, b, as it is similar for the other experiments. We know that α_{rp} is a Nash equilibrium of the estimated game analyzed in Fig. 4a (meta Table not shown). The outcome of α_{rp}



(a) Trajectory plot for α_{rp}, α_{vp}, and α_{rv}

(b) Trajectory plot for α_{rvp}, α_{vp}, and α_{rv}

Fig. 4 AlphaGo evolutionary dynamics plots for $\alpha_{rp}, \alpha_{vp}, \alpha_{rv}$, and α_{rvp}

against α_{rv} was estimated with $n_{\alpha_{rp},\alpha_{rv}} = 63$ games (for the other pair of strategies we have $n_{\alpha_{vp},\alpha_{rp}} = 65$ and $n_{\alpha_{vp},\alpha_{rv}} = 133$). Because of the symmetry of the problem, the bound in Sect. 5.2.1 is reduced to:

$$1 - \delta = 1 - \exp(\log(2p|S|) - 2\epsilon^2 n_{\min})$$

Therefore, we can conclude that the strategy α_{rp} is an 2ϵ-Nash equilibrium (with $\epsilon = 0.15$) for the real game with probability at least 0.29 .

The same calculation would also give a confidence of 0.35 for the RD studied in Fig. 4b for an $\epsilon = 0.15$ (as the number of samples are $(n_{\alpha_{rv},\alpha_{vp}}, n_{\alpha_{vp},\alpha_{rvp}}, n_{\alpha_{rvp},\alpha_{rv}}) = (65, 106, 91)$). For these two cases, $\epsilon = 0.05$ is too small to give any form of guarantee in probability. For example, the number of samples necessary per joint strategy to provide an accurate estimation for $\epsilon = 0.05$ with a confidence of $1 - \delta = 0.95$ would be at least 1097 samples (819 samples for $(\epsilon, \delta) = (0.05, 0.2)$ and 122 samples for $(\epsilon, \delta) = (0.15, 0.05)$).

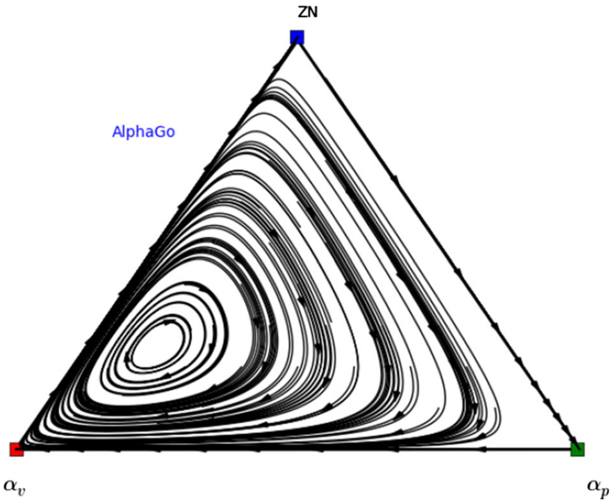


Fig. 5 Intransitive behaviour for α_v, α_p, and Zen

6.1.3 Experiment 3: Cyclic behaviour

A final experiment investigates what happens if we add a *pre-AlphaGo* state-of-the-art algorithm to the strategy space. We have observed that even though α_{rvp} remains the strongest strategy, beating all other *AlphaGo* versions and previous state-of-the-art algorithms, cyclic behaviour can occur, something that cannot be measured or seen from Elo ratings.[2] More precisely, we constructed a meta-game payoff table for strategies α_v, α_p and Zen (one of the previous commercial state-of-the-art algorithms). In Fig. 5 we have plotted the evolutionary dynamics for this meta-game, and as can be observed there is a mixed equilibrium in strategy space, around which the dynamics cycle, indicating that Zen is capable of introducing in-transitivity, as α_v beats α_p, α_p beats Zen and Zen beats α_v.

6.2 Colonel Blotto

Colonel Blotto is a resource allocation game originally introduced by Borel [4]. Two players interact, each allocating m troops over n locations. They do this separately without communication, after which both distributions are compared to determine the winner. When a player has more troops in a specific location, it wins that location. The player winning the most locations wins the game. This game has many game theoretic intricacies, for an analysis see [17]. Kohli et al. have run Colonel Blotto on Facebook (project Waterloo), collecting data describing how humans play this game, with each player having $m = 100$ troops and considering $n = 5$ battlefields. The number of strategies in the game is vast: a game with m troops and n locations has $\binom{m+n-1}{n-1}$ strategies.

Based on Kohli et al. we carry out a meta game analysis of the *strongest strategies* and the *most frequently played strategies* on Facebook (here the meta-game analysis is simply a restricted game). We have a look at several 3-strategy simplexes, which can be considered as 2-faces of the entire strategy space.

[2] An Elo rating or score is a measure to express the relative strength of a player, or strategy [8]. It was named after Arpad Elo and originally introduced to rate chess players. For an introduction see e.g. [7].

Table 8 5 of the strongest strategies played on Facebook

Strongest strategies		
Strategy	Frequency	Win rate
[36, 35, 24, 3, 2]	1	.74
[37, 37, 21, 3, 2]	17	.73
[35, 35, 26, 2, 2]	1	.73
[35, 34, 25, 3, 3]	3	.70
[35, 35, 24, 3, 3]	13	.70

Table 9 Meta-game payoff table generated for strategies $s_1 = [36, 35, 24, 3, 2]$, $s_2 = [37, 37, 21, 3, 2]$, and $s_3 = [35, 35, 26, 2, 2]$

s_1	s_2	s_3	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	0.66	0	0.34
0	2	0	0	0.5	0
1	1	0	0.33	0.67	0
0	0	2	0	0	0.5
0	1	1	0	0.75	0.25

An instance of a strategy in the game of Blotto will be denoted as follows: $[t_1, t_2, t_3, t_4, t_5]$ with $\sum_i t_i = 100$. All permutations σ_i in this division of troops belong to the same strategy. We assume that permutations are chosen uniformly by a player. Note that in this game there is no need to carry out the theoretical analysis of the approximation of the meta-game, as we are are not examining heuristics or strategies over Blotto strategies, but rather these strategies themselves, for which the payoff against any other strategy will always be the same (by computation). Nevertheless, carrying out a meta-game analysis reveals interesting information.

6.2.1 Experiment 1: Top performing strategies

In this first experiment we examine the dynamics of the simplex consisting of the three best scoring strategies from the study of [17]: [36, 35, 24, 3, 2], [37, 37, 21, 3, 2], and [35, 35, 26, 2, 2], see Table 8. In a first step we compute a meta-game payoff table for these three strategies. The interactions are pairwise, and the expected payoff can be easily computed, assuming a uniform distribution for different permutations of a strategy. This normalised payoff is shown in Table 9.

Using Table 9 we can compute evolutionary dynamics using the standard replicator equation. The resulting trajectory plot can be observed in Fig. 6a.

The first thing we see is that we have one strong attractor, i.e, strategy $s_2 = [37, 37, 21, 3, 2]$ and there is transitive behaviour, meaning that [36, 35, 24, 3, 2] beats [35, 35, 26, 2, 2], [37, 37, 21, 3, 2] beats [36, 35, 24, 3, 2], and [37, 37, 21, 3, 2] beats [35, 35, 26, 2, 2]. Although [37, 37, 21, 3, 2] is the strongest strategy in this 3-strategy meta-game, the win rates (computed over all played strategies in project Waterloo) indicate that strategy [36, 35, 24, 3, 2] was more successful on Facebook. The differences are minimal, and on average it is better to choose [37, 37, 21, 3, 2], which was also the most frequently chosen strategy from the set of strong strategies, see Table 8. We show a similar plot for the evolutionary dynamics of strategies [35, 34, 25, 3, 3], [37, 37, 21, 3, 2], and [35, 35, 24, 3, 3] in Fig. 6b, which are three of the most frequently played strong strategies from Table 8.

Fig. 6 **a** Dynamics of [36, 35, 24, 3, 2], [37, 37, 21, 3, 2], and [35, 35, 26, 2, 2]. **b** Dynamics of [35, 34, 25, 3, 3], [37, 37, 21, 3, 2], and [35, 35, 24, 3, 3]

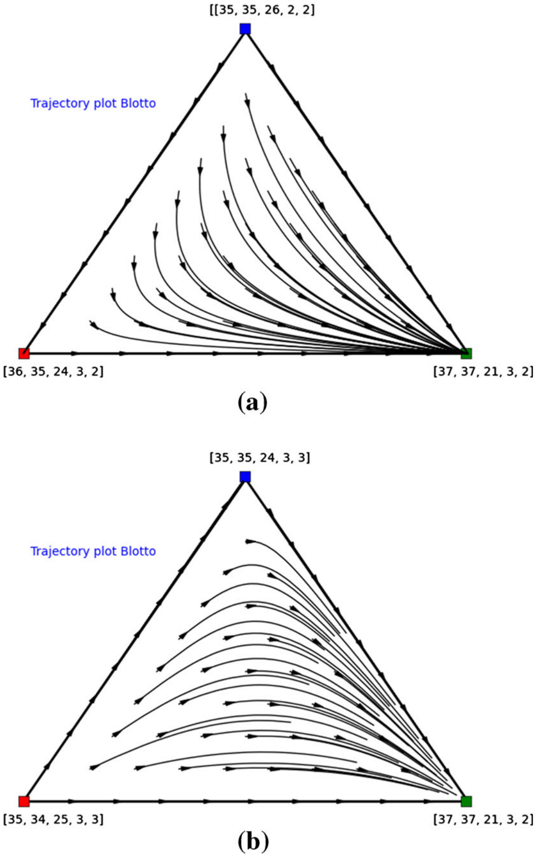


Table 10 The 8 most frequently played strategies on Facebook

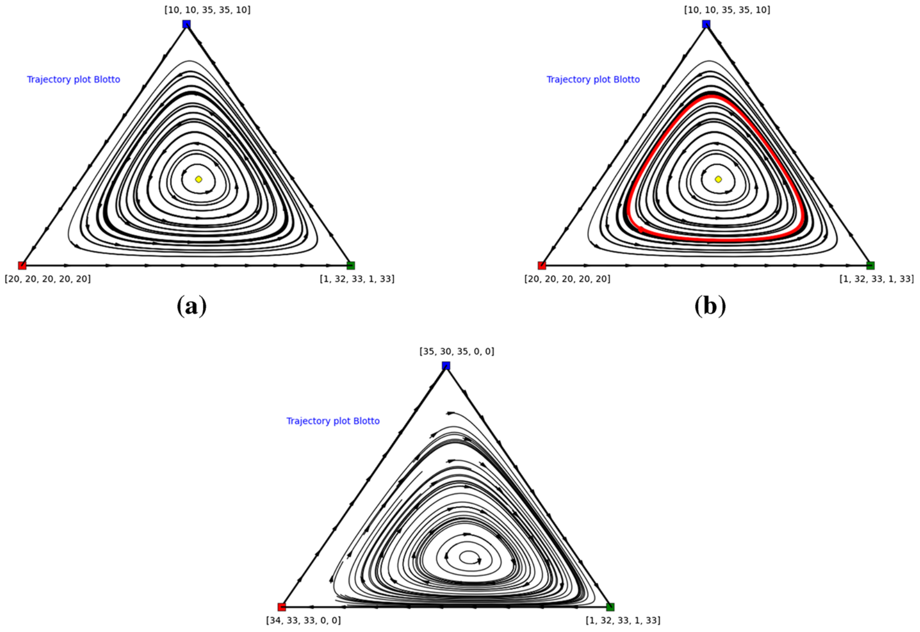
Most played strategies	
Strategy	Frequency
[34, 33, 33, 0, 0]	271
[20, 20, 20, 20, 20]	235
[33, 1, 33, 0, 33]	127
[1, 32, 33, 1, 33]	97
[35, 30, 35, 0, 0]	68
[0, 100, 0, 0, 0]	67
[10, 10, 35, 35, 10]	58
[25, 25, 25, 25, 0]	50

6.2.2 Experiment 2: Most frequently played strategies

We compared the evolutionary dynamics of the eight most frequently played strategies and present here a selection of some of the results. The meta-game under study in this domain concerns a 2-type repeated NFG G with $|S| = 8$. We will look at various 2-faces of the 8-simplex. The top eight most frequently played strategies are shown in Table 10.

Table 11 Meta-game payoff table generated for strategies $s_1 = [20, 20, 20, 20, 20]$, $s_2 = [1, 32, 33, 1, 33]$, and $s_3 = [10, 10, 35, 35, 10]$

s_1	s_2	s_3	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	1	0	0
0	2	0	0	0.5	0
1	1	0	0	1	0
0	0	2	0	0	0.5
0	1	1	0	0.1	0.9

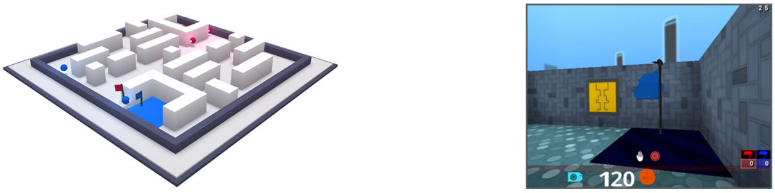


(a) **(b)** **(c)**

Fig. 7 Dynamics of 3 2-faces of the 8-simplex: **a** Nash eq. **b** Human play, **c** another example of intransitive behaviour

First we investigate the strategies [20, 20, 20, 20, 20], [1, 32, 33, 1, 33], and [10, 10, 35, 35, 10] from our strategy set. In Table 11 we show the resulting meta-game payoff table of this 2-face simplex. Using this table we can again compute the replicator dynamics and investigate the trajectory plots in Fig. 7a. We observe that the dynamics cycle around a mixed Nash equilibrium (every interior rest point is a Nash equilibrium). This intransitive behaviour makes sense by looking at the pairwise interactions between strategies and the corresponding payoffs they receive from Table 9. The expected payoff for [20, 20, 20, 20, 20], when playing against [1, 32, 33, 1, 33] will be lower than the expected payoff for [1, 32, 33, 1, 33]. Similarly, [1, 32, 33, 1, 33] will be beaten by [10, 10, 35, 35, 10] when they meet, and to make the cycle complete, [10, 10, 35, 35, 10] will receive a lower expected payoff against [20, 20, 20, 20, 20]. As such, the behaviour will cycle around a the Nash equilibrium.

An interesting question is where human players are situated in this cyclic behaviour landscape. In Fig. 7b we show the same trajectory plot but added a red marker to indicate the strategy profile based on the frequencies of these 3 strategies played by human players. This is derived from Table 10 and the profile vector is (0.6, 0.25, 0.15). If we assume that



(a) Example procedurally generated map. **(b)** Example first-person agent observation.

Fig. 8 Capture the Flag: teams of two agents (shown as blue and red spheres), must pick up the opposing team's flag and return it to their own base. **a** Shows the blue agent capturing the red team's flag, and **b** is the blue agent's first-person observation where they can see their own flag in their base

Table 12 Meta-game payoff table generated for strategies α_{130}, α_{170}, and α_{90}

a_{130}	a_{170}	a_{90}	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	0.88	0	0.12
0	2	0	0	0.5	0
1	1	0	0.22	0.78	0
0	0	2	0	0	0.5
0	1	1	0	0.72	0.28

the human agents optimise their behaviour in a *survival of the fittest* style they will cycle along the red trajectory. In Fig. 7c we illustrate similar intransitive behaviour for three other frequently played strategies.

6.3 Capture the flag

CTF is a game involving multiple players competing to capture each other's flags. We specifically consider the implementation introduced by Jaderberg et al. [13] featuring two opposing teams consisting of two players which compete to capture each other's flags by strategically navigating, tagging, and evading opponents. Matches between two teams take place in a procedurally generated map, and the team with the greatest number of flag captures within five minutes is determined as the winner. In Fig. 8, we present an example map which the agents play in, as well as an example first-person observation that the agents see.

The data set we study consists of the FTW strategy from [13] (which achieved human-level performance in CTF) at various points in its training, as well as a set of rule-based strategies across a number of skill levels. Teams are formed of two strategies, and for each team we have a number of matches against each other team from which we can summarise the wins and losses between strategies meeting several times. In the following experiments we restrict each team to two of the same strategy, and henceforth refer to teams by their strategy.

6.3.1 Experiment 1: Strategies throughout training

In our first experiment, we examine the FTW strategy from [13] at various points in its training, i.e. at training step $90e7$, $130e7$, and $170e7$ (referred to as α_{90}, α_{130}, and α_{170} respectively). To begin, we compute a meta-game payoff table for these strategies, with the normalised payoff shown in Table 12.

Using Table 12, we can compute evolutionary dynamics using the standard replication equation, with the resulting trajectory plot presented in Fig. 9a. As can be seen, there is one

Table 13 Meta-game payoff table generated for strategies α_{130}, α_{170}, and $Tauri$

a_{130}	a_{170}	$Tauri$	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	0.57	0	0.43
0	2	0	0	0.5	0
1	1	0	0.28	0.72	0
0	0	2	0	0	0.5
0	1	1	0	0.41	0.59

strong attractor (i.e. α_{170}), as well as transitive behaviour where α_{170} beats α_{130}, α_{170} beats α_{90}, and α_{130} beats α_{90}.

We can now examine how good of an approximation the estimated game is. From the RD studied in Fig. 9a, we know that α_{170} is a Nash equilibrium of the estimated game. Using the formulation from Sect. 6.1.2, we can conclude that the strategy α_{170} is a 2ϵ-Nash equilibrium (with ϵ=0.15) for the real game with probability at least 0.99, where the number of samples are $(n_{\alpha_{90},\alpha_{130}}, n_{\alpha_{90},\alpha_{170}}, n_{\alpha_{130},\alpha_{170}}) = (164, 174, 157)$. And again here, $\epsilon = 0.05$ is too small to provide guarantees with that few samples.

6.3.2 Experiment 2: Cyclic behavior

We now investigate what happens if we add a rule-based algorithm to the strategy space (referred to as $Tauri$). To begin, we construct a meta-game payoff table for strategies α_{130}, α_{170}, and $Tauri$ as shown in Table 13. We can then plot the evolutionary dynamics for this meta-game. From the RD in Fig. 9b, we can see there is a mixed equilibrium in the strategy space around which the dynamics cycle, indicating that $Tauri$ is capable of introducing in-transitivity as α_{170} beats α_{130}, α_{130} beats $Tauri$, and $Tauri$ beats α_{170}.

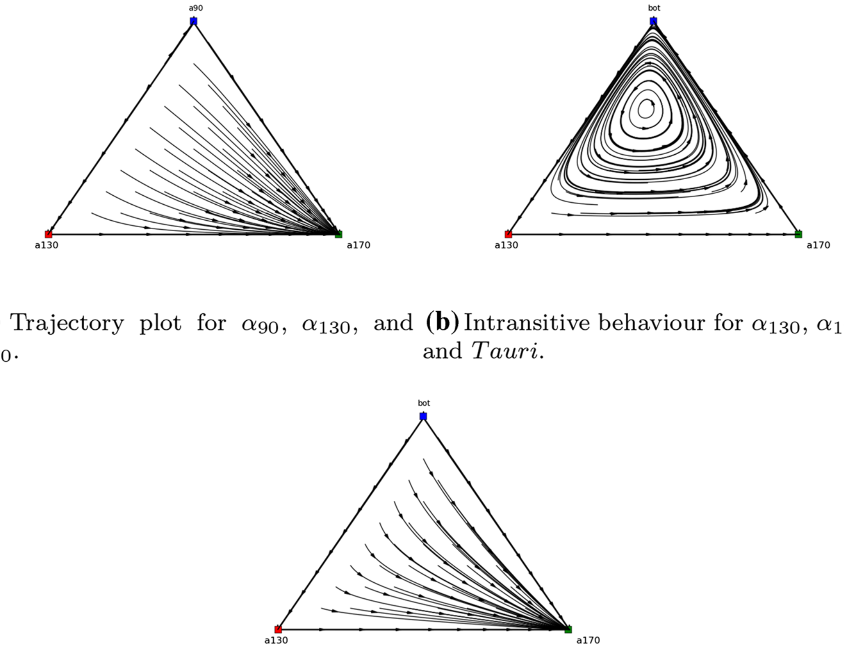
As before, we can now examine how good of an approximation the estimated game is. With the samples of $(n_{\alpha_{130},\alpha_{170}}, n_{\alpha_{170},\alpha_{Tauri}}, n_{\alpha_{130},\alpha_{Tauri}}) = (157, 80, 96)$, we can conclude that the strategy α_{170} is a 2ϵ-Nash equilibrium (with $\epsilon = 0.15$) for the real game with probability at least 0.67 (We can't guarantee anything with $\epsilon = 0.05$).

Figure 12 in the supplemental material additionally illustrates the mixed strategy dynamics using Boltzmann Q-learning [33] dynamics for the three pairwise combinations of $(\alpha_{130}, \alpha_{170}, Tauri)$.

6.3.3 Experiment 3: Elo scores

In Experiment 2, we saw that while α_{170} remains the strongest strategy even with the addition of $Tauri$, cyclic behaviour can now occur. This is a behavior which cannot be measured or seen from Elo ratings. To show this, we first calculate the Elo ratings of all strategies. We do this by iterating through the full data set of matches between the strategies severl times, updating the Elo rating of each strategy based on the outcome of each match until they converge.

Using this method, we determine the strategies α_{130}, α_{170}, and $Tauri$ as having an Elo rating of 1546, 1648, and 1520 respectively. We can then compute the probability of s_i beating s_j using $p(s_i) = \frac{1}{(1+10^m)}$, where m is the rating difference between s_i and s_j divided by 400. We summarize the win rates of these three strategies in Table 14, and show the resulting computed evolutionary dynamics in Fig. 9c. As can be observed, there is no cyclic behavior and α_{170} is the Nash equilibrium of the estimated game. In other words, the cyclic behavior of the three strategies cannot be seen from Elo ratings as it assumes transitivity by design.



(a) Trajectory plot for α_{90}, α_{130}, and α_{170}.

(b) Intransitive behaviour for α_{130}, α_{170}, and $Tauri$.

(c) Trajectory plot for α_{130}, α_{170}, and $Tauri$.

Fig. 9 Dynamics of 3 2-faces of the 8-simplex: **a** throughout training. **b** Intransitive behaviour by addition of $Tauri$. **c** Using Elo-calculated win probabilities

Table 14 Win rates calculated from Elo ratings for strategies α_{130}, α_{170}, and $Tauri$

a_{130}	a_{170}	$Tauri$	U_{i1}	U_{i2}	U_{i3}
2	0	0	0.5	0	0
1	0	1	0.54	0	0.46
0	2	0	0	0.5	0
1	1	0	0.36	0.64	0
0	0	2	0	0	0.5
0	1	1	0	0.68	0.32

6.4 PSRO-generated meta-game

We now turn our attention to an asymmetric game. Policy Space Response Oracles (PSRO) is a multiagent reinforcement learning process that reduces the strategy space of large extensive-form games via iterative best response computation. PSRO can be seen as a generalized form of fictitious play that produces approximate best responses, with arbitrary distributions over generated responses computed by meta-strategy solvers. One application of PSRO was applied to a commonly-used benchmark problem known as Leduc poker [28], except with a fixed action space and penalties for taking illegal moves. Therefore PSRO learned to play from scratch, without knowing which moves were legal. Leduc poker has a deck of 6 cards (jack, queen, king in two suits). Each player receives an initial private card, can bet a fixed amount of 2 chips in the first round, 4 chips in the second round, with a maximum of two raises in each round. A public card is revealed before the second round starts.

Table 15 Asymmetric PSRO meta game applied to Leduc poker

	D	E	F
A	−2.26, 0.02	−2.06, −1.72	−1.65, −1.43
B	−4.77, −0.13	−4.02, −3.54	−5.96, −2.30
C	−2.71, −1.77	−2.52, −2.94	−6.10, 1.06

Table 16 Left: First counterpart game of the PSRO empirical game. Right: Second counterpart game of the PSRO empirical game

	A	B	C		D	E	F
A	−2.26	−2.06	−1.65	D	0.02	−1.72	−1.43
B	−4.77	−4.02	−5.96	E	−0.13	−3.54	−2.30
C	−2.71	−2.52	−6.10	F	−1.77	−2.94	1.06

In Table 15 we present such an asymmetric 3 × 3 2-player game generated by the first few epochs of PSRO learning to play Leduc Poker. In the game illustrated here, each player has three strategies that, for ease of the exposition, we call $\{A, B, C\}$ for player 1, and $\{D, E, F\}$ for player 2. Each one of these strategies represents an approximate best response to a distribution over previous opponent strategies. . In Table 16 we show the two symmetric counterpart games (see Sect. 4.3) of the empirical game produced by PSRO.

Again we can now analyse the equilibrium landscape of this game, but now using the asymmetric meta-game payoff table and the decomposition result introduced in Sect. 4.3. Since the PSRO meta game is asymmetric we need two populations for the asymmetric replicator equations. Analysing and plotting the evolutionary asymmetric replicator dynamics now quickly becomes very tedious as we deal with two simplices, one for each player. More precisely, if we consider a strategy profile for one player in its corresponding simplex, and that player is adjusting its strategy, this will immediately cause the second simplex to change, and vice versa. Consequently, it is not straightforward anymore to analyse the dynamics.

In order to facilitate the process of analysing the dynamics we can apply the counterpart theorems to remedy the problem. In Figs. 10 and 11 we show the evolutionary dynamics of the counterpart games. As can be observed in Fig. 10 the first counterpart game has only one equilibrium, i.e., a pure Nash equilibrium in which both players play strategy A, which absorbs the entire strategy space. Looking at Fig. 11 we see the situation is a bit more complex in the second counterpart game, here we observe three equilibiria: one pure at strategy D, one pure at strategy F, and one unstable mixed equilibrium at the 1-face formed by strategies D and F. All these equilibria are Nash in the respective counterpart games[3]. By applying the theory of Sect. 4.3 we now know that we only maintain the combination $((1, 0, 0), (1, 0, 0))$ as a pure Nash equilibrium of the asymmetric PSRO empirical game, since these strategies have the same support as a Nash equilibrium in the counterpart games. The other equilibria in the second counterpart game can be discarded as candidates for Nash equilibria in the PSRO empirical game since they do not appear as equilibria for player 1.

Finally, each joint action of the game was estimated with 100 samples. As the outcome of the game is bounded in the interval $[-13, 13]$ we can only guarantee that the Nash equilibrium of the meta game we studied is a 2ϵ-Nash equilibrium of the unknown underlying game. It

[3] Banach solver (http://banach.lse.ac.uk/) is used to check Nash equilibria [1].

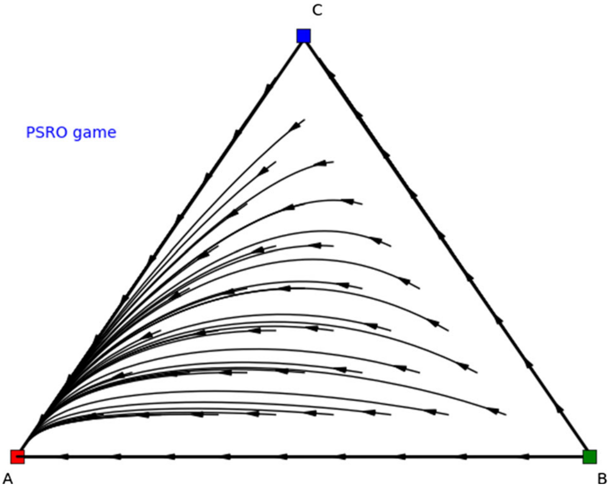


Fig. 10 Trajectory plot of the first CP game

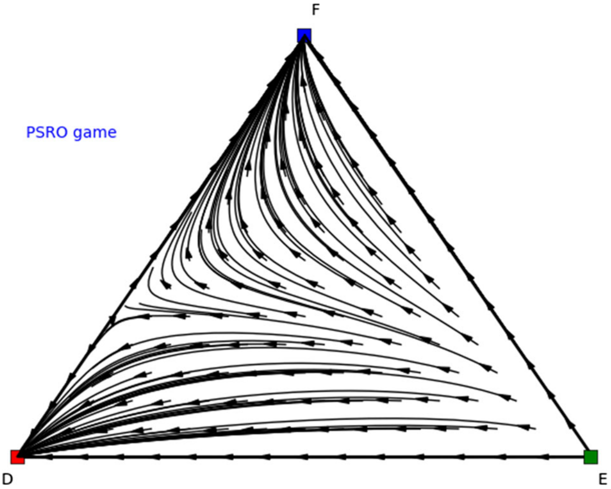


Fig. 11 Trajectory plot of the 2nd CP game

turns out that with $n = 100$ and $\epsilon = 0.05$, the confidence can only be guaranteed to be above 10^{-8}. To guarantee a confidence of at least 0.95 for the same value of $\epsilon = 0.05$, we would need at least $n = 886 \times 10^3$ samples.

7 Conclusion

In this paper we have provided some bounds for empirical game theoretic analysis using the heuristic payoff table method introduced by Walsh et al. [39] for both symmetric and 2-player asymmetric games. We call such games *meta-games* as they consider complex strategies instead of atomic actions as found in normal-form games. As such they are well suited to investigate real-world multi-agent interactions, as they summarize behaviour in terms of high-level strategies rather than primitive actions. We use the fact that a Nash equilibrium of the meta-game is a 2ϵ-Nash equilibrium of the true underlying game to provide theoretical bounds on how much data samples are required to build a reliable meta payoff table. As such our method allows for an equilibrium analysis with a certain confidence that this game is a good approximation of the underlying meta game. Finally, we have carried out an empirical illustration of this method in four complex domains, i.e., *AlphaGo*, Colonel Blotto, CTF and PSRO, showing the feasibility and strengths of the approach.

Acknowledgements We wish to thank Angeliki Lazaridou, Jean-Baptiste Lespiau, Guy Lever and Michael Wellman for insightful comments, the DeepMind AlphaGo team for support with the analysis of the AlphaGo dataset, and Pushmeet Kohli for supporting us with the Colonel Blotto dataset. This paper is an extended version of our AAMAS 2018 article [31]. It contains more background descriptions, improved theoretical bounds, provides additional experiments and includes some supplemental material on theoretical properties of heuristic payoff tables.

Appendix: Supplemental material

A Theoretical properties of heuristic payoff tables

This section introduces some theoretical properties of HPTs that have not been closely examined before, and which are useful as further background to understand some of the experimental results in a wider game theoretic context. Furthermore, these are not typically discussed in the EGTA literature and as such can be useful for newcomers to the area. This section can also be skipped as these results are not necessary to understand the main theoretical results of this paper.

A.1 Equivalence with mixed strategy dynamics

Lemma 1 *Let the number of players $p = 2$. Then infinite population replicator dynamics is equivalently the evolutionary dynamics of mixed strategies, as defined by the replicator equation.*

Proof Denote the population composition by $C = (\alpha_1, \ldots \alpha_k)$, with $0 \leq \alpha_i \leq 1$ and $\sum_i \alpha_i = 1$. Equivalently view this as a mixed strategy M. Let E_i be the expected payoff to strategy i when 2-player matches are sampled according to the population composition C. Let $\tilde{E}_i$ be the expected payoff to strategy i when played in a single 2-player match against the mixed strategy M. It suffices to show that $E_i = \tilde{E}_i$.

Without loss of generality, we may assume that i is played by player 1. Indeed, either the game is symmetric, in which case i can be swapped to the first slot, or it is asymmetric, in which case the dynamics operate on prescribed slots. Then the payoff E_i is the average payoff of i against the pure strategies in the distribution given by C, which is identical to the payoff $\tilde{E}_i$ of the i against the mixed strategy M, by definition. Since the population is infinite, M is an arbitrary discrete distribution across strategies, completing the proof. □

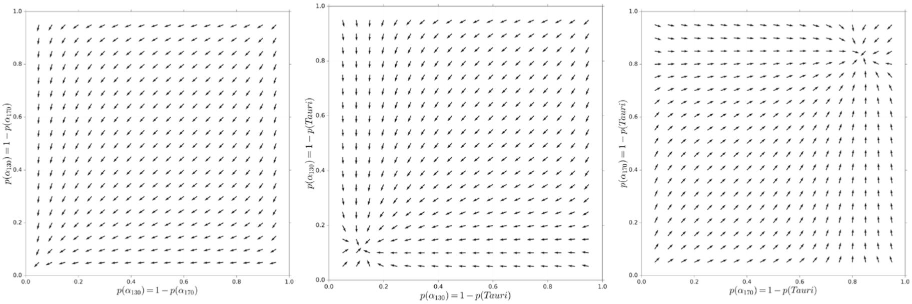


Fig. 12 2-dimensional replicator dynamics for Capture the Flag with Boltzmann Q-learning dynamics [33]. The pure strategy α_{170} dominates all mixes with α_{130}, as one might expect given that the latter strategy has benefited from longer training time. The equilibrium when playing the learned strategies against the bot are mixed, with a low weight on α_{130} and a high weight on α_{170}

Hence, the replicator dynamics of mixed strategies in a 2-player normal form is well approximated by the replicator dynamics of a sufficiently large population of pure strategies playing sufficiently many games according to the appropriate population composition.

A.2 Example: 2-dimensional plots for Capture The Flag

The results in A.1 have consequences for visualization. In particular, when $n = k = 2$ we can plot the mixed strategy dynamics in 2 dimensions, since they are completely determined by the probability of each player choosing the first strategy. More generally one could use the asymmetric replicator dynamics for n-players to prove a similar equivalence result to Lemma 1 and hence plot mixed strategy dynamics in this case. However, the dimensionality of this data is $n^k - n$, which limits the applicability of such a visualization.

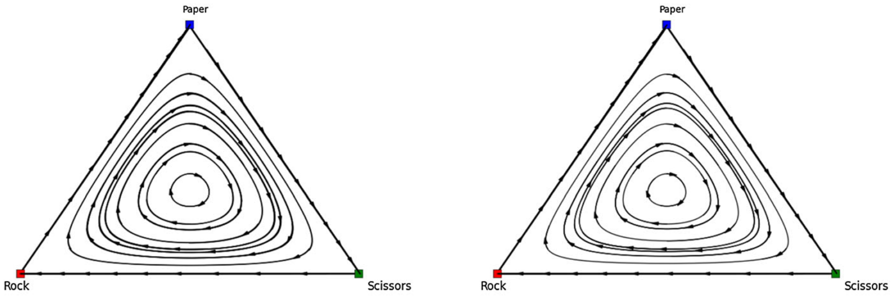
As an example we show in Fig. 12 the directional field plots for the three pairwise combinations of the strategies $(\alpha_{130}, \alpha_{170}, Tauri)$ in the Capture the Flag dataset discussed in Sect. 6.3. This gives us further insight into the equilibria possible if agents were allowed to mix their strategies in the meta-game. Indeed, such an analysis is important, since human players are certainly adept at mixing different styles of play across repeated interactions.

A.3 Approximation of zero-sum NFGs by HPTs

In principle small populations of agents (often $O(k)$, where k denotes the number of strategies) do not generate a good approximation of the replicator dynamics of the underlying game. Here we show that in the case of 2-player symmetric zero-sum games, the case we primarily consider in this paper, that the approximation is a good one and one case use equally well HPTs as NFGs.

Lemma 2 *In a symmetric zero-sum 2-player game the single-population replicator dynamics with population size 2 applied to the HPT representation of the game is the same as the replicator dynamics of the true underlying game modulo a factor* $\alpha = \frac{2}{2-x_i}$.

Proof We compute an approximation to $(Ax)_i$ using the HPT methodology outlined in Sect. 4, and use the same notation as in that section. First we write $U_{ij} = W_{pqj}$ with $p \leq q$ where p and q denote the strategies that each player plays given the distribution N_i. Then, since the game is zero-sum, $W_{pqj} = A_{jq}$ when $j = p \neq q$, $W_{pqj} = A_{jp}$ when $j = q \neq p$ and $W_{pqj} = 0$ otherwise.



(a) Evolutionary dynamics of the Rock-Paper-Scissors game using the NFG representation.

(b) Evolutionary dynamics of the Rock-Paper-Scissors game using the HPT representation.

Fig. 13 Evolutionary dynamics of Rock-Paper-Scissors using the NFG and HPT representations

Now we must calculate $P(N_i|x)$ for those i for which U_{ij} is not identically zero. By the above observation such i yield N_i with support on two distinct strategies, so the multinomial coefficient is $\frac{2!}{1!1!} = 2$. Therefore we may write

$$r_j(x) = \frac{2\sum_{p,q,p<q} W_{pqj} x_p x_q}{1-(1-x_j)^2} = \frac{2\sum_q A_{jq} x_j x_q}{x_j(2-x_j)} = \frac{2}{2-x_j}(Ax)_j \,. \tag{3}$$

□

Note that one can also use win probabilities for a zero-sum game. The same results holds and the derivation of the lemma is analogous.

Corollary 1 *In a zero-sum* 2*-player game, the replicator dynamics with population size* 2 *are a good approximation of the replicator dynamics of the true underlying game.*

Proof Since A is an antisymmetric matrix, the true replicator dynamics reduce to

$$\dot{x_i} = x_i(Ax)_i \,. \tag{4}$$

The replicator dynamics derived from the approximate expected payoffs are

$$\dot{x_i} = x_i \left(\frac{2(Ax)_i}{2-x_i} - \sum_{jk} \frac{2x_j A_{jk} x_k}{(2-x_j)} \right) \tag{5}$$

$$= x_i \left((Ax)_i + \frac{x_i(Ax)_i}{2-x_i} - \sum_{jk} \frac{x_j^2 A_{jk} x_k}{(2-x_j)} \right) . \tag{6}$$

We go from Eqs. 5 to 6 by using $\frac{2}{(2-x_i)} = 1 + \frac{x_i}{(2-x_i)}$. The first term in Eq. 6 gives the true dynamics. The second term is a homogeneous scaling of the first, so it has no significant effect other than to change the evolution rate of the replicators. The third term could in principle change the dynamics in a non-trivial way. When $x_i \gg x_j$ for $j \neq i$, the third term is dominated by the first two. When $x_j \gg x_i$ for $j \neq i$ the whole expression is very small, due to the x_i prefactor. Therefore, for the final term to contribute maximally, we would need $x_j \approx x_i \approx \frac{1}{2}$ for some j. But then the final term is down by a factor of at least $\left(\frac{1}{2}\right)^2 \frac{2}{3} = \frac{1}{6}$ compared with the first, which means it can be ignored. □

An example illustrates the corollary. In Fig. 13 we show the dynamics of Rock-Paper-Scissors using both the exact NFG and the HPT with population size 2. One can observe that they are almost identical.

References

1. Avis, D., Rosenberg, G., Savani, R., & von Stengel, B. (2010). Enumeration of nash equilibria for two-player games. *Economic Theory*, *42*, 9–37.
2. Bloembergen, D., Hennes, D., McBurney, P., & Tuyls, K. (2015). Trading in markets with noisy information: An evolutionary analysis. *Connection Science*, *27*, 253–268.
3. Bloembergen, D., Tuyls, K., Hennes, D., & Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research (JAIR)*, *53*, 659–697.
4. Borel, E. (1921). La théorie du jeu les équations intégrales à noyau symétrique. *comptes rendus de l'académie*, *173*, 1304–1308. english translation by savage, l.: The theory of play and integral equations with skew symmetric kernels. Econometrica **21**, 97–100 (1953).
5. Brinkman, E., & Wellman, M. (2016). Shading and efficiency in limit-order markets. In *Proceedings of the IJCAI-16 workshop on algorithmic game theory*.
6. Cassell, B. A., & Wellman, M. (2013). EGTAOnline: An experiment manager for simulation-based game studies. In F. Giardini & F. Amblard (Eds.), *Multi-Agent-Based Simulation XIII, Lecture Notes in Computer Science* (Vol. 7838). Berlin: Springer.
7. Coulom, R. (2008). Whole-history rating: A Bayesian rating system for players of time-varying strength. In *Computers and games, 6th international conference*, CG 2008, Beijing, China, September 29–October 1, 2008. Proceedings (pp. 113–124).
8. Elo, A. E. (1978). *The rating of chess players, past and present*. Bronx: Ishi Press International.
9. Erev, I., & Roth, A. E. (2007). Multi-agent learning and the descriptive value of simple models. *Artificial Intelligence*, *171*(7), 423–428.
10. Gintis, H. (2009). *Game theory evolving* (2nd ed.). Princeton, NJ: University Press.
11. Hennes, D., Claes, D., & Tuyls, K. (2013). Evolutionary advantage of reciprocity in collision avoidance. In *Proceedings of the AAMAS 2013 workshop on autonomous robots and multirobot systems* (ARMS 2013)
12. Hofbauer, J., & Sigmund, K. (1998). *Evolutionary games and population dynamics*. Cambridge: Cambridge University Press.
13. Jaderberg, M., Czarnecki, W. M., Dunning, I., Marris, L., Lever, G., Castaneda, A. G., Beattie, C., Rabinowitz, N. C., Morcos, A. S., Ruderman, A., et al. (2018). *Human-level performance in first-person multiplayer games with population-based deep reinforcement learning*. arXiv preprint arXiv:1807.01281.
14. Jecmen, S., Brinkman, E., & Sinha, A. (2018). Bounding regret in simulated games. In: ICML workshop on exploration in RL.
15. Julian Schvartzman, L., & Wellman, M. P. (2009). Stronger CDA strategies through empirical game-theoretic analysis and reinforcement learning. *AAMAS*, *1*, 249–256.
16. Kaisers, M., Tuyls, K., Thuijsman, F., & Parsons, S. (2008). Auction analysis by normal form game approximation. In *Proceedings of the 2008 IEEE/WIC/ACM international conference on intelligent agent technology*, Sydney, NSW, Australia, December 9–12, 2008 (pp. 447–450).
17. Kohli, P., Kearns, M., Bachrach, Y., Herbrich, R., Stillwell, D., & Graepel, T. (2012). Colonel blotto on facebook: the effect of social relations on strategic interaction. In *Web science 2012, WebSci '12, Evanston*, IL, USA, June 22–24, 2012 (pp. 141–150).
18. Lanctot, M., Zambaldi, V., Gruslys, A., Lazaridou, A., Tuyls, K., Perolat, J., et al. (2017). A unified game-theoretic approach to multiagent reinforcement learning. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4190–4203). Berlin: Springer.
19. Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflicts. *Nature*, *246*, 15–18.
20. Nguyen, T., Wright, M., Wellman, M., & Singh, S. (2017). Multi-stage attack graph security games: Heuristic strategies, with empirical game-theoretic analysis. In *Proceedings of the fourth ACM workshop on moving target defense*.
21. Nijssen, P., & Winands, M. H. (2012). Monte carlo tree search for the hide-and-seek game scotland yard. *IEEE Transactions on Computational Intelligence and AI in Games*, *4*(4), 282–294.
22. Phelps, S., Cai, K., McBurney, P., Niu, J., Parsons, S., & Sklar, E. (2007). Auctions, evolution, and multi-agent learning. In *Adaptive agents and multi-agent systems III. Adaptation and multi-agent learning, 5th,*

6th, and 7th European symposium, ALAMAS 2005-2007 on adaptive and learning agents and multi-agent systems, Revised Selected Papers (pp. 188–210).

23. Phelps, S., Parsons, S., & McBurney, P. (2004). An evolutionary game-theoretic comparison of two double-auction market designs. In *Agent-mediated electronic commerce VI, theories for and engineering of distributed mechanisms and systems, AAMAS 2004 Workshop*, AMEC 2004, New York, NY, USA, July 19, 2004, Revised Selected Papers (pp. 101–114).
24. Ponsen, M., Tuyls, K., Kaisers, M., & Ramon, J. (2009). An evolutionary game-theoretic analysis of poker strategies. *Entertainment Computing*, *1*(1), 39–45.
25. Prakash, A., & Wellman, M. (2015). Empirical game-theoretic analysis for moving target defense. In *Proceedings of the second ACM workshop on moving target defense*
26. Rosenfeld, A., & Kraus, S. (2018). *Predicting human decision-making: From prediction to action. Synthesis lectures on artificial intelligence and machine learning*. San Rafael: Morgan & Claypool Publishers.
27. Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., et al. (2016). Mastering the game of go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.
28. Southey, F., Bowling, M., Larson, B., Piccione, C., Burch, N., Billings, D., & Rayner, C. (2005). Bayes' bluff: Opponent modelling in poker. In *Proceedings of the twenty-first conference on uncertainty in artificial intelligence (UAI-05)*.
29. Stone, P., & Veloso, M. (1998). Towards collaborative and adversarial learning: A case study in robotic soccer. *International Journal of Human-Computer Studies*, *48*(1), 83–104.
30. Tuyls, K., & Parsons, S. (2007). What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, *171*(7), 406–416.
31. Tuyls, K., Pérolat, J., Lanctot, M., Leibo, J. Z., & Graepel, T. (2018). A generalised method for empirical game theoretic analysis. In *International foundation for autonomous agents and multiagent systems (AAMAS)*, Richland, SC, USA/ACM (pp. 77–85).
32. Tuyls, K., Perolat, J., Lanctot, M., Savani, R., Leibo, J., Ord, T., et al. (2018). Symmetric decomposition of asymmetric games. *Scientific Reports*, *8*(1), 1015.
33. Tuyls, K., Verbeeck, K., & Lenaerts, T. (2003). A selection-mutation model for q-learning in multi-agent systems. In *The second international joint conference on autonomous agents & multiagent systems, AAMAS 2003*, July 14–18, 2003, Melbourne, Victoria, Australia, Proceedings (pp. 693–700).
34. Vorobeychik, Y. (2010). Probabilistic analysis of simulation-based games. *ACM Transactions on Modeling and Computer Simulation*, *20*(3), 16. https://doi.org/10.1145/1842713.1842719.
35. Vorobeychik, Y., & Wellman, M. P. (2008). Stochastic search methods for nash equilibrium approximation in simulation-based games. In *Proceedings of the seventh international conference on autonomous agents and multiagent systems (AAMAS)* (pp. 1055–1062).
36. Vorobeychik, Y., Wellman, M. P., & Singh, S. (2007). Learning payoff functions in infinite games. *Machine Learning*, *67*, 145–168.
37. Wah, E., Hurd, D., & Wellman, M. (2015). Strategic market choice: Frequent call markets vs. continuous double auctions for fast and slow traders. In *Proceedings of the third EAI conference on auctions, market mechanisms, and their applications*.
38. Wah, E., Wright, M., & Wellman, M. (2017). Welfare effects of market making in continuous double auctions. *Journal of Artificial Intelligence Research*, *59*, 613–650.
39. Walsh, W. E., Das, R., Tesauro, G., & Kephart, J. (2002). Analyzing complex strategic interactions in multi-agent games. In *AAAI-02 workshop on game theoretic and decision theoretic agents, 2002*.
40. Walsh, W. E., Parkes, D. C., & Das, R. (2003). Choosing samples to compute heuristic-strategy nash equilibrium. In *Proceedings of the fifth workshop on agent-mediated electronic commerce*.
41. Wang, X., Vorobeychik, Y., & Wellman, M. (2018). A cloaking mechanism to mitigate market manipulation. In *Proceedings of the 27th international joint conference on artificial intelligence* (pp. 541–547).
42. Weibull, J. (1997). *Evolutionary game theory*. Cambridge: MIT Press.
43. Wellman, M., Kim, T., & Duong, Q. (2013). Analyzing incentives for protocol compliance in complex domains: A case study of introduction-based routing. In *Proceedings of the 12th workshop on the economics of information security*.
44. Wellman, M. P. (2006). Methods for empirical game-theoretic analysis. In *Proceedings, The twenty-first national conference on artificial intelligence and the eighteenth innovative applications of artificial intelligence conference* July 16–20, 2006, Boston, Massachusetts, USA (pp. 1552–1556).
45. Wiedenbeck, B., Cassell, B. A., & Wellman, M. P. (2014). Bootstrap statistics for empirical games. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 597–604). International Foundation for Autonomous Agents and Multiagent Systems.
46. Wiedenbeck, B., & Wellman, M. (2012). Scaling simulation-based game analysis through deviation-preserving reduction. In *Proceedings of the eleventh international conference on autonomous agents and multiagent systems (AAMAS)*.

47. Wright, M. (2016). Using reinforcement learning to validate empirical game-theoretic analysis: A continuous double auction study. CoRR arXiv:1604.06710.
48. Wright, M., Venkatesan, S., Albenese, M., & Wellman, M. (2016). Moving target defense against DDoS attacks: An empirical game-theoretic analysis. In *Proceedings of the third ACM workshop on moving target defense*.
49. Wright, M., & Wellman, M. P. (2018). Evaluating the stability of non-adaptive trading in continuous double auctions. *AAMAS*, 614–622.
50. Zeeman, E. C., (1980). Population dynamics from game theory. In *Global theory of dynamical systems*, Springer, Berlin, Heidelberg (pp. 471–497).
51. Zeeman, E. (1981). Dynamics of the evolution of animal conflicts. *Theoretical Biology*, *89*, 249–270.