

On Altruism and Spite in Bimatrix Games

Michail Fasoulakis¹, Leonidas Bakopoulos², Charilaos Akasiadis², and Georgios Chalkiadakis²

¹ Royal Holloway, University of London

michail.fasoulakis@rhul.ac.uk

² Technical University of Crete

{lbakopoulos,cakasiadis,ghalkiadakis}@tuc.gr

Abstract. One common assumption in game theory is that any player optimizes a utility function that takes into account only its own payoff. However, it has long been observed that in real life players may adopt an altruistic or even spiteful behaviour. As such, there are numerous attempts in the economics literature that strive to explain the fact that players are not entirely selfish, but most of these works do not focus on the algorithmic implications of altruism or spite in games. In this paper, we relax the aforementioned “self-interest” assumption, and initiate the study of algorithmic aspects of bimatrix games—such as the complexity and the quality of their (approximate) Nash equilibria—under altruism or spite. We provide both a theoretical and an experimental treatment of these topics. Moreover, we demonstrate the potential for learning the degree of an opponent’s altruistic/spiteful behaviour, and employing this for opponent selection and transfer of knowledge in bimatrix games.

Keywords: Altruism · Spite · Approximate · Nash Equilibria · Bimatrix games

1 Introduction

One assumption widely used in game theory and economics, is the so-called Self-Interest Hypothesis (SIH), according to which players/agents³ aim to maximize their personal payoff. However, SIH validity in the real world has been questioned. Experiments in the 1980s and 1990s have demonstrated that people are not, in general, self-interested. As a result, several models attempt to explain this phenomenon by introducing concepts such as reciprocity, fairness, and altruism into decision-making processes [26, 24, 29, 17].

As the seminal work of [29] points out, while several different models have been developed to describe and organize the evidence of non-selfish behaviour, there is no general model that provides a complete description of observational findings. Several models put forward specific functional forms for *interdependent preferences*, assuming that individuals seek to optimize well-defined preferences which are nevertheless dependent on the behaviour of others; and then proceed to

³ We will be using the terms *players* and *agents* interchangeably in this paper.

make predictions of agents' strategies based on the equilibrium behaviour ensuing from the utility functions describing the preferences in question. The notion of *spite*, which corresponds to various degrees of malicious or envious behaviour exhibited by individuals, also emerges in various manifestations given the different functional forms of preferences and utility defined in such models [21, 7, 29].

In our work in this paper, we take a step towards an in-depth algorithmic study of economic models departing from the SIH hypothesis. We focus on bimatrix games, one of the most fundamental classes of games that can potentially give us insights to the study of games with more than two players.

In such games we examine (both from a theoretical and from an experimental point of view) the computational complexity of equilibria arising when selfishness is relaxed and the concepts of *altruism* and *spitefulness* (or *spite*) are introduced. Note that by "selfish" we refer to rational agents whose utility is a function of their own payoff alone, i.e., theirs is a "typical" utility function; while altruistic or spiteful agents essentially maximize a utility function that also incorporates others' utility.

More specifically, our contributions are as follows. We adopt the concept of *altruism* from [23] as was introduced for n -player games, and proceed to define *spite* in two-player (bimatrix) games in a straightforward manner. We define a game in which players may exhibit a behaviour that is taking into account the utility of the opponents—i.e., may exhibit a level of altruism or spite. First, we prove that in almost any case of altruism/spite, the computation of an exact NE remains PPAD-complete, and then we study approximate Nash equilibria (NE) under different levels of altruism/spite, providing polynomial-time algorithms for specific fundamental cases.

Interestingly, for specific levels of altruism or spite, not only can we find *in polynomial-time* an ε -NE that is *better* than the state-of-the-art approximation for polynomial-time algorithms for approximate Nash equilibria [14], but we also show this has *maximum social welfare*.

In addition, we design an algorithmic framework and test altruism/spite in numerous bimatrix games. We put forward a gradient descent-based algorithm that identifies strategies along with an optimal setting of the spite/altruism parameters, such that an ε -equilibrium with low ε is found.

We test our algorithm experimentally for computing an approximate NE in games considering altruism/spite.

Moreover, we provide experiments to showcase the *learning* of the λ values corresponding to the *unknown levels of altruism/spite* of opponents; and thus allow rational agents to use learned values to beneficially "select" which opponent to face in a bimatrix game. In addition, we demonstrate the potential for the transfer of knowledge regarding the opponents' levels of altruism/spite, and for "transfer learning" [25] among different games. As such, our work takes a first step in combining algorithmic game theory with machine learning within this framework.

2 Preliminaries

Basic game-theoretic concepts. We consider bimatrix games⁴ $G = (R, C) \in [0, 1]^{n \times n}$, where R is the payoff matrix of the row player (the “Row”) and C the payoff matrix of the column player (the “Column”). Any player has n pure strategies at her disposal and any pure strategy i is denoted as the column vector e_i , which has 1 in the index i and zero elsewhere. A player can also play mixed strategies, as probability distributions $\mathbf{x} \in \Delta^{n-1}$ on her pure strategies, where Δ^{n-1} is the $n - 1$ dimensional simplex created by her pure strategies. A pair of mixed strategies (\mathbf{x}, \mathbf{y}) is called a strategy profile, where \mathbf{x}, \mathbf{y} are the mixed strategies of the Row and Column, respectively.

Let (\mathbf{x}, \mathbf{y}) be a strategy profile, then the payoff for the Row is $\mathbf{x}^T R \mathbf{y}$ and the payoff for the Column is $\mathbf{x}^T C \mathbf{y}$. In this paper, we consider that the players are expected payoff maximizers, in other words, they play strategies to maximize their expected payoff, given the strategy of the other player. A strategy profile is a social optimum in the utilitarian sense (sum of the payoffs) if it is the maximum over all strategy profiles. We can easily see that in bimatrix games there is at least one pure strategy profile that is a social optimum, and that it can be computed in polynomial-time by exhaustive search over the pure strategy profiles.

For any strategy profile (\mathbf{x}, \mathbf{y}) we define as *regret* or *approximation* of this strategy profile for the Row player the function:

$$\begin{aligned} f_R(\mathbf{x}, \mathbf{y}) : \Delta^{n-1} \times \Delta^{n-1} &\rightarrow [0, 1], \text{ with} \\ f_R(\mathbf{x}, \mathbf{y}) &= \max_i e_i^T R \mathbf{y} - \mathbf{x}^T R \mathbf{y}, \end{aligned}$$

and, similarly, for the Column the function:

$$\begin{aligned} f_C(\mathbf{x}, \mathbf{y}) : \Delta^{n-1} \times \Delta^{n-1} &\rightarrow [0, 1], \text{ with} \\ f_C(\mathbf{x}, \mathbf{y}) &= \max_j \mathbf{x}^T C e_j - \mathbf{x}^T C \mathbf{y}. \end{aligned}$$

The (approximate) Nash equilibria are defined as follows.

Definition 1 (Nash equilibria (NE)). A strategy (\mathbf{x}, \mathbf{y}) is a Nash equilibrium, if and only if, for any i, j , it holds that:

$$\mathbf{x}^T R \mathbf{y} \geq e_i^T R \mathbf{y}, \text{ and } \mathbf{x}^T C \mathbf{y} \geq \mathbf{x}^T C e_j,$$

or, in other words $f_R(\mathbf{x}, \mathbf{y}) = 0$ and $f_C(\mathbf{x}, \mathbf{y}) = 0$ (the regrets of both players are zero). Furthermore, for approximate Nash equilibria, we have the following definition.

⁴ Any bimatrix game can be transformed to a bimatrix game of entries in $[0, 1]$ with exactly the same Nash equilibria.

Definition 2 (ε -approximate Nash equilibria). A strategy (\mathbf{x}, \mathbf{y}) is an ε -approximate Nash equilibrium, for any $\varepsilon \in [0, 1]$, if and only if, for any i, j , it holds that:

$$\mathbf{x}^T R \mathbf{y} + \varepsilon \geq \mathbf{e}_i^T R \mathbf{y}, \text{ and } \mathbf{x}^T C \mathbf{y} + \varepsilon \geq \mathbf{x}^T C \mathbf{e}_j,$$

or in other words $f_R(\mathbf{x}, \mathbf{y}) \leq \varepsilon$ and $f_C(\mathbf{x}, \mathbf{y}) \leq \varepsilon$ (the regrets of both players are at most ε). It is easy to see that we have an exact NE if and only if $\varepsilon = 0$.

Altruism and spite. In this paper, we consider the level of altruism or spite that players introduce as a policy/behaviour to an initial game (R, C) , with two parameters $\lambda_R, \lambda_C \in [-1, 1]$, for the row and Column, respectively. The game is thus transformed to the following modified bimatrix game:

$$G' = (R', C') = \left(R + \lambda_R \cdot C, C + \lambda_C \cdot R \right).$$

In particular, we consider the level of altruism for strictly positive values of λ 's as in [23].

Definition 3 (Altruism [23]). The Row is altruistic of a level λ_R , if and only if, $\lambda_R > 0$. Similar for the Column.

In the same spirit, here we introduce and define the level of spite (or spitefulness) for strictly negative values.

Definition 4 (Spite). The Row is spiteful of a level $|\lambda_R|$, if and only if, $\lambda_R < 0$. Similar for the Column.

The remaining case of λ 's being 0, is the case where $G' = (R', C') = (R, C) = G$, i.e., the initial and the modified games coincide.

Note that an agent's strategy depends on the policy/behavior (altruistic, spiteful, or SIH-compliant) that is “prescribed” to her by her λ , which drives her to maximize her expected payoff in the modified game G' .

3 Game theoretic aspects of altruism and spite

We now proceed to make some initial observations for the modified payoff game in a game theoretic context. In the case of altruism the payoff matrix of the Row is:

$$R' = R + \lambda_R \cdot C = (1 - \lambda_R) \cdot R + \lambda_R \cdot (C + R),$$

similar for the Column. In this case, the payoff matrix is a convex combination between the own payoff matrix (selfish interest) and the social optimum (altruistic interest), where the latter is the utilitarian optimum (the sum of payoffs).

In the case of spite the Row's payoff matrix is:

$$R' = (1 + \lambda_R) \cdot R - \lambda_R \cdot (R - C).$$

Interestingly, this is a convex combination⁵ between the own payoff matrix (selfish interest) and a matrix with payoff $R - C$ (spiteful interest), where the Row maximizes the difference of the payoffs between her own payoff and the payoff of the other player. Similar for the Column.

Another interesting observation is that in the case of equivalent policy/behaviour (both same level of altruism, spite, or selfish), in other words $\lambda = \lambda_R = \lambda_C$, we have:

$$\begin{aligned} G' &= (R + \lambda \cdot C, C + \lambda \cdot R) \\ &= (1 - \lambda) \cdot \frac{(R - C, C - R)}{2} + (1 + \lambda) \cdot \frac{(R + C, R + C)}{2}. \end{aligned}$$

We see that if $\lambda = -1$ the game is zero-sum, and if $\lambda = 1$ it is a potential game/common payoff—both poly-time solvable. In any other case, we have a general-sum game. This interesting result can be seen as a complement to the well-known result that any bimatrix game (i.e., one with $\lambda = 0$ in our context) is the average of one zero-sum and one potential game [19]. Finally, if Row is altruistic with $\lambda > 0$, and Column spiteful by the same level λ (or symmetrically), then G' becomes:

$$\begin{aligned} G' &= (R + \lambda \cdot C, C - \lambda \cdot R) \\ &= (1 - \lambda) \cdot (R, C) + \lambda \cdot (R + C, C - R). \end{aligned}$$

In this case, we see that the modified game is a convex combination of the initial selfish interest one, and the game where one player maximizes the social optimum, while the other the difference between hers and the opponent's payoff.

4 Computational complexity of computing a Nash Equilibrium under altruism or spite

We know that computing a NE in bimatrix games is PPAD-complete [10, 13] when both players seek to maximize their own payoff only. On the other hand, if players have an altruistic or spiteful behaviour, the Nash equilibria of the modified game may change and in some cases be completely different. Thus, we are interested whether this change can affect the computational complexity of the problem. We study this under the presence of different combinations and levels of altruistic/spiteful players' behaviours.

It is easy to see that in the extreme case of $\lambda_R = \lambda_C = 1$ (both players have the maximum level of altruism), a strategy profile that maximizes the utilitarian optimum (sum of payoffs) of the game G is an NE in the game G' . This strategy profile can be easily found in polynomial time by exhaustive search on the pure strategy profiles of the game. Furthermore, in the other extreme case of $\lambda_R = \lambda_C = -1$ (both players have the maximum level of spite), the game G'

⁵ Note that λ_R is a negative number.

is a zero-sum game, so we can compute again in polynomial time a NE by linear programming [1]. For almost any remaining case⁶, we have the following Theorem.

Theorem 1. *For any $\lambda_R \in (-1, 1)$ and any $\lambda_C \in (-1, 1)$, the problem of computing a NE in G' is PPAD-complete.*

Proof. Let an arbitrary bimatrix game $(A, B) \in [0, 1]^{n \times n}$. As mentioned above, by [10], we know that the problem of computing a NE in a bimatrix game is PPAD-complete. Let now the game

$$(R, C) = (A - \lambda_R \cdot B, B - \lambda_C \cdot A),$$

then we have that

$$\begin{aligned} G' &= \left(A - \lambda_R \cdot B + \lambda_R \cdot B - \lambda_R \cdot \lambda_C \cdot A, \right. \\ &\quad \left. B - \lambda_C \cdot A + \lambda_C \cdot A - \lambda_C \cdot \lambda_R \cdot B \right) \\ &= \left((1 - \lambda_R \cdot \lambda_C) \cdot A, (1 - \lambda_R \cdot \lambda_C) \cdot B \right), \\ &= (1 - \lambda_R \cdot \lambda_C) \cdot (A, B), \end{aligned}$$

which we can easily prove, since $(1 - \lambda_R \cdot \lambda_C) > 0$, that any NE of this game is also a NE in the game (A, B) .

Note that the previous result does not imply anything about the complexity of the initial game (R, C) from which the reduction starts. It does imply, however, that in general if we introduce altruism/spite, the computation of a Nash equilibrium in the modified game is PPAD-complete.

Having these negatives results, assuming that $P \neq \text{PPAD}$, we are interested in the effect of altruism/spite in the computation of approximate Nash equilibria in our context as we see in the following section.

5 Regret under altruism or spite

With the previous complexity results for the exact Nash equilibria at hand, we now focus on the case of approximate NE in our context. First, we study some fundamental cases where the players can indeed achieve approximate Nash equilibria in polynomial-time with low regret under altruism/spite, and then we give an algorithmic framework as a “learning” algorithm where the players in an alternating optimization fashion can compute their behavior/policy in order to achieve approximate Nash equilibria with good regret. Our initial results are the following.

⁶ We leave open the case of $\lambda_R \cdot \lambda_C = -1$.

Theorem 2. Let a bimatrix game $G = (R, C) \in [0, 1]^{n \times n}$ and the modified game $G' = (R + \lambda \cdot C, C + \lambda \cdot R) \in [0, 1]^{n \times n}$, with $\lambda > 0$ (Both players are altruistic of the same level). Then, the social optimum (utilitarian optimum) of the game G' is an $(1 - \lambda)$ -Nash equilibrium.

Proof. Let (e_i, e_j) be the social optimum (Utilitarian optimum) of the game G' . Then, it holds that, for any k, ℓ :

$$\begin{aligned} (1 + \lambda) \cdot R_{ij} + (1 + \lambda) \cdot C_{ij} \\ \geq (1 + \lambda) \cdot R_{k\ell} + (1 + \lambda) \cdot C_{k\ell}. \end{aligned}$$

Consider now the case that $\ell = j$. Since, $1 + \lambda > 0$, the previous inequality implies that

$$\begin{aligned} R_{ij} + C_{ij} &\geq R_{kj} + C_{kj}, \\ &\rightarrow \\ R_{ij} + C_{ij} - (1 - \lambda) \cdot C_{ij} - (1 - \lambda) \cdot C_{kj} &\geq R_{kj} + C_{kj} \\ &\quad - (1 - \lambda) \cdot C_{ij} - (1 - \lambda) \cdot C_{kj}, \\ &\rightarrow \\ R_{ij} + \lambda \cdot C_{ij} - (1 - \lambda) \cdot C_{kj} &\geq R_{kj} + \lambda \cdot C_{kj} \\ &\quad - (1 - \lambda) \cdot C_{ij}, \\ &\rightarrow \\ R_{ij} + \lambda \cdot C_{ij} + (1 - \lambda) \cdot (C_{ij} - C_{kj}) &\geq R_{kj} + \lambda \cdot C_{kj}. \end{aligned}$$

Since $C \in [0, 1]^{n \times n}$, then $C_{ij} - C_{kj} \leq 1$. This, by definition, implies that (e_i, e_j) has regret at most $(1 - \lambda)$. Similar, for the other player.

Note that in the case of study of the game G' in the previous Theorem, the Utilitarian optimum of the game G' is also an optimum for the game G . Thus, our Theorem implies that we can in poly-time find a strategy profile with $(1 - \lambda)$ approximation and also maximum social welfare, even for the initial game. Furthermore, in the case of level of altruism $\lambda > \frac{2}{3}$, not only we can find in polynomial-time an ε -NE with $\varepsilon < \frac{1}{3}$ that is better than the state of art approximation for polynomial-time algorithms for approximate Nash equilibria [14], but also this has maximum social welfare in the game. We now give a similar result for the case of spite as follows:

Theorem 3. Let a bimatrix game $G = (R, C) \in [0, 1]^{n \times n}$ and the modified game $G' = (R + \lambda \cdot C, C + \lambda \cdot R) \in [0, 1]^{n \times n}$, with $\lambda < 0$ (Both players are spiteful of the same level). Then, there is a polynomial time to compute an $(1 + \lambda)$ -Nash equilibrium.

Proof. We can analyse the game G' as $G' = (R - C, C - R) + (1 + \lambda) \cdot (C, R)$. For the zero-sum game $(R - C, C - R)$ we can find in polynomial time a Nash

equilibrium (x, y) by Linear programming [1]. Then, it holds by the definition of the NE that $x^T(R - C)y \geq e_i^T(R - C)y$, for any i . Thus, we have that:

$$\begin{aligned} x^T(R + \lambda \cdot C)y &= x^T(R - C)y + (1 + \lambda) \cdot x^T Cy \\ &\geq e_i^T(R - C)y + (1 + \lambda) \cdot x^T Cy, \\ &\quad \rightarrow \\ x^T(R + \lambda \cdot C)y + e_i^T Cy - (1 + \lambda) \cdot x^T Cy &\geq e_i^T Ry, \\ &\quad \rightarrow \\ x^T(R + \lambda \cdot C)y + e_i^T Cy + \lambda \cdot e_i^T Cy - (1 + \lambda) \cdot x^T Cy &\geq e_i^T Ry + \lambda \cdot e_i^T Cy, \\ &\quad \rightarrow \\ x^T(R + \lambda \cdot C)y + (1 + \lambda) \cdot e_i^T Cy &\geq e_i^T(R + \lambda \cdot C)y. \end{aligned}$$

The last inequality holds since $(1 + \lambda) \cdot x^T Cy \geq 0$. Thus, we have that $(1 + \lambda) \cdot e_i^T Cy \leq 1 + \lambda$, since $e_i^T Cy \leq 1$, and this implies that the regret of the row player is at most $(1 + \lambda)$. Similar for the column player.

Similarly as above, we can see that in the case of level of spite of $\lambda < -\frac{2}{3}$ (high spite), we can find in polynomial-time an ε -NE, with $\varepsilon < \frac{1}{3}$, better than the state of art approximation algorithms of Nash equilibria in bimatrix games [14].

For other cases of altruism/spite we will introduce an algorithmic framework of study not only the regret, but also the optimal policies of the player to converge to approximate NE with good regret. This framework is based on the idea of the descent method of the paper of [30] to achieve approximate NE in bimatrix games.

This algorithm is based on performing gradient descent to minimize the maximum of the two regrets function (we will refer to this as *the TS function* for now), a method proposed in [30] for computing a 0.3393-NE. The high level idea is that the descent method converges to a stationary point, and then, given its properties, a more sophisticated strategy profile can be constructed, so that this profile or the stationary one have no more than 0.3393 regret.

Here, we only focus on the descent method and the regrets of the stationary point that are reached. But before we introduce our framework, we give some simple illustrative examples that show the possible change of the stationary points under altruism/spite that inspired our approach.

As proven in [11], there is a tight example/game with a stationary point of the TS function of approximation 0.3393. Inspired by this, we provide and study the following similar example as:

$$R = \begin{pmatrix} 0.1 & 0 & 0 \\ 0.1 + b & 1 & 1 \\ 0.1 + b & 0 & 0 \end{pmatrix}, C = \begin{pmatrix} 0.1 & 0.1 + b & 0.1 + b \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad (1)$$

with b equal to 0.3393. Note that in this game the strategy profile $x_s = y_s = [1 \ 0 \ 0]$ is a stationary point with 0.3393 approximation and the point $x^* =$

$y^* = [0 \ 1 \ 0]$ is an exact Nash equilibrium. Let us now consider the case of introducing a level of altruism/spite with λ 's in this game. The modified game $G' = (R', C')$ will be as follows:

$$R' = \begin{pmatrix} 0.1 + \lambda_R \cdot 0.1 & \lambda_R \cdot (0.1 + b) & \lambda_R \cdot (0.1 + b) \\ 0.1 + b & 1 + \lambda_R & 1 \\ 0.1 + b & \lambda_R & 0 \end{pmatrix},$$

$$C' = \begin{pmatrix} 0.1 + \lambda_C \cdot 0.1 & 0.1 + b & 0.1 + b \\ \lambda_C \cdot (0.1 + b) & 1 + \lambda_C & \lambda_C \\ \lambda_C \cdot (0.1 + b) & 1 & 0 \end{pmatrix}.$$

When $\lambda_R = 0$ and $\lambda_C = \frac{1}{2}$, the modified game⁷ becomes

$$R' = \begin{pmatrix} 0.1 & 0 & 0 \\ 0.1 + b & 1 & 1 \\ 0.1 + b & 0 & 0 \end{pmatrix}, C' = \begin{pmatrix} 0.15 & 0.1 + b & 0.1 + b \\ \frac{0.1+b}{2} & 1.5 & \frac{1}{2} \\ \frac{0.1+b}{2} & 1 & 0 \end{pmatrix}.$$

We can see that, since the regrets in the modified game at $(\mathbf{x}_s, \mathbf{y}_s)$ are not anymore equal, this strategy profile cannot be a stationary point in the new game for the TS objective function, since it has been proven that any stationary point must have the same regrets (see e.g. Lemma 2.4 in [16]). This example shows us that the stationary points can be changed under altruism or spite.

However, this is not always the case. E.g., consider the following game:

$$R = \begin{pmatrix} 0 & 0 & 0 \\ b & 1 & 1 \\ b & 0 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & b & b \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad (2)$$

In this game $x_s = y_s = [1 \ 0 \ 0]$ is also a stationary point. Let now the modified game with $\lambda_R = 0, \lambda_C = 1/2$, as

$$R' = \begin{pmatrix} 0 & \lambda_R \cdot b & \lambda_R \cdot b \\ b & 1 + \lambda_R & 1 \\ b & \lambda_R & 0 \end{pmatrix}, C' = \begin{pmatrix} 0 & b & b \\ \lambda_C \cdot b & 1 + \lambda_C & \lambda_C \\ \lambda_C \cdot b & 1 & 0 \end{pmatrix}.$$

Then, $\mathbf{x}_s = \mathbf{y}_s = [1 \ 0 \ 0]$ remains a stationary point. Inspired by these observations, we provide an algorithmic framework based on gradient-based approaches to achieve optimal behavior/policies for better approximation of the Nash equilibria for the strategy profile of the players.

Our Algorithmic framework. Motivated by the example above, we now propose an algorithmic framework in which players start the game with a “selfish” behaviour (λ 's = 0), and subsequently change it—or, in other words, *learn a*

⁷ Here for simplicity of the presentation we don't scale the payoffs of the players to $[0, 1]$.

different policy/ behaviour—so as to achieve *a better*, in terms of regret in the modified game, strategy profile.

In particular, Algorithm 1 constitutes an alternating process that optimizes, in terms of regret, the strategy profiles $(\mathbf{x}^t, \mathbf{y}^t)$ given the level of altruism/spite $(\lambda_R^{t-1}, \lambda_C^{t-1})$, and vice versa. Specifically, starting from $(\lambda_R^0, \lambda_C^0) = (0, 0)$, we apply the projected gradient descent [5] (PGD) procedure on a quadratic program with fixed values of λ^{t-1} 's, having a particular objective regarding the regrets, for instance minimizing the maximum of the regrets as indicated in program (3) below,⁸ to converge to stationary point strategies $(\mathbf{x}^t, \mathbf{y}^t)$. Then, having as fixed the strategy profile $(\mathbf{x}^t, \mathbf{y}^t)$, the stationary point of the previous step, we find the optimal policies $(\lambda_R^t, \lambda_C^t)$, that minimizes the objective function, by solving a linear program, LP (5). This procedure is repeated for a maximum of h times, or as the algorithm converges.

Algorithm 1 An alternating optimization process to produce a regret-minimizing $\langle(\mathbf{x}^t, \mathbf{y}^t), (\lambda_R^t, \lambda_C^t)\rangle$ pair

```

1: Input: A bimatrix game  $(R, C) \in [0, 1]^{n \times n}$ , an initial strategy profile  $(\mathbf{x}^0, \mathbf{y}^0)$  and  $(\lambda_R^0, \lambda_C^0)$ .
2: for  $t = 1$  to  $h$  do
3:   // Find a stationary point  $(\mathbf{x}^t, \mathbf{y}^t)$  given  $(\lambda_R^{t-1}, \lambda_C^{t-1})$ 
4:    $(\mathbf{x}^t, \mathbf{y}^t) \leftarrow$  Solution of the quadratic program (3) using PGD with  $\lambda_R = \lambda_R^{t-1}$  and  $\lambda_C = \lambda_C^{t-1}$ .
5:   // Find optimal  $(\lambda_R^t, \lambda_C^t)$  given strat. profile  $(\mathbf{x}^t, \mathbf{y}^t)$ 
6:    $(\lambda_R^t, \lambda_C^t) \leftarrow$  Solution of the linear program (5) with  $\mathbf{x} = \mathbf{x}^t$  and  $\mathbf{y} = \mathbf{y}^t$ .
7: end for
8: Return  $(\mathbf{x}^{t-1}, \mathbf{y}^{t-1})$  and  $(\lambda_R^{t-1}, \lambda_C^{t-1})$ 
```

In more detail, with respect to the first part of the strategy profiles optimization given λ 's, we adopt a quadratic formulation of (3) inspired by [22, 30, 14]:

$$\begin{aligned}
& \min_{\mathbf{x}, \mathbf{y}, \pi_1, \pi_2} \max \left\{ \pi_1 - \mathbf{x}^T R \mathbf{y} - \lambda_R \mathbf{x}^T C \mathbf{y}, \right. \\
& \quad \left. \pi_2 - \mathbf{x}^T C \mathbf{y} - \lambda_C \mathbf{x}^T R \mathbf{y} \right\} \\
\text{s.t. } & \mathbf{e}_i^T (R + \lambda_R C) \mathbf{y} \leq \pi_1, \text{ for any } i \in [n], \\
& \mathbf{x}^T (C + \lambda_C R) \mathbf{e}_j \leq \pi_2, \text{ for any } j \in [n], \\
& \sum_i \mathbf{x}_i = 1; \sum_j \mathbf{y}_j = 1; \pi_1, \pi_2 \in [-1, 1].
\end{aligned} \tag{3}$$

Note that in [22] a similar quadratic formulation was proposed:

⁸ Alternatively, the sum of the regrets, as in program 4.

$$\begin{aligned}
& \min_{\mathbf{x}, \mathbf{y}, \pi_1, \pi_2} \pi_1 - \mathbf{x}^T R \mathbf{y} - \lambda_R \mathbf{x}^T C \mathbf{y} \\
& \quad + \pi_2 - \mathbf{x}^T C \mathbf{y} - \lambda_C \mathbf{x}^T R \mathbf{y} \\
\text{subject to } & \mathbf{e}_i^T (R + \lambda_R C) \mathbf{y} \leq \pi_1, \text{ for any } i \in [n], \\
& \mathbf{x}^T (C + \lambda_C R) \mathbf{e}_j \leq \pi_2, \text{ for any } j \in [n], \\
& \sum_i \mathbf{x}_i = 1; \sum_j \mathbf{y}_j = 1; \pi_1, \pi_2 \in [-1, 1].
\end{aligned} \tag{4}$$

In the above quadratic programs, given altruism/spite levels (λ_R, λ_C) of the players, the result (\mathbf{x}, \mathbf{y}) is a stationary point for the modified game. We note here that the only exact Nash equilibria of the game are the global optima, but the stationary points that are different than the global optima are nevertheless ε -approximate Nash equilibria, with $\varepsilon > 0$. It is well-known that such a formulation can lead to a stationary point that is not necessarily a Nash equilibrium.

On the other hand, given a strategy profile we solve the following linear program (5) for finding the optimal behaviour for minimizing the maximum of the regrets:

$$\begin{aligned}
& \min_{(\lambda_R, \lambda_C)} \gamma \\
\text{s.t. } & \gamma \geq \mathbf{e}_i^T (R + \lambda_R C) \mathbf{y} - \mathbf{x}^T (R + \lambda_R C) \mathbf{y}, \\
& \gamma \geq \mathbf{x}^T (C + \lambda_C R) \mathbf{e}_j - \mathbf{x}^T (C + \lambda_C R) \mathbf{y}, \\
& \text{for any } i, j \in [n]; \lambda_R, \lambda_C \in [-1, 1].
\end{aligned} \tag{5}$$

The following is the LP finding the optimal behavior for minimizing the sum of the regrets function:

$$\begin{aligned}
& \min_{(\lambda_R, \lambda_C)} \gamma \\
\text{s.t. } & \gamma \geq \mathbf{e}_i^T (R + \lambda_R C) \mathbf{y} - \mathbf{x}^T (R + \lambda_R C) \mathbf{y} \\
& + \mathbf{x}^T (C + \lambda_C R) \mathbf{e}_j - \mathbf{x}^T (C + \lambda_C R) \mathbf{y}, \\
& \text{for any } i, j \in [n]; \lambda_R, \lambda_C \in [-1, 1].
\end{aligned} \tag{6}$$

Analysis of Algorithm 1. Theorem 4 shows the convergence of Algorithm 1:

Theorem 4. Let a bimatrix game $G = (R, C) \in [0, 1]^{n \times n}$ and r^t be the regret function⁹ of the two players in the modified game $G' = (R + \lambda_R^t \cdot C, C + \lambda_C^t \cdot R) \in [0, 1]^{n \times n}$, for any $t \geq 0$. Then, it holds that $r^{t+1} < r^t$, for any t s.t. $(\mathbf{x}^t, \mathbf{y}^t)$ is not a stationary point given $(\lambda_R^t, \lambda_C^t)$, and $(\lambda_R^t, \lambda_C^t)$ is not an optimal pair given $(\mathbf{x}^t, \mathbf{y}^t)$.

Proof. Let an iteration t of the algorithm, with $(\mathbf{x}^t, \mathbf{y}^t)$ be a strategy profile, $(\lambda_R^t, \lambda_C^t)$ be the current a pair of λ 's and r^t be the value if the regret function in

⁹ Maximum regret, or sum of the regrets.

the modified game in this strategy profile. The PGD algorithm will find a new stationary point $(\mathbf{x}_s, \mathbf{y}_s)$, minimizing the maximum of the regrets, s.t. $r_s < r^t$ by definition of the quadratic programming (3), since $(\mathbf{x}^t, \mathbf{y}^t)$ is not optimal given $(\lambda_R^t, \lambda_C^t)$. Let now that we fix this profile, then the LP (5), by definition, will find a new pair of λ 's, $(\lambda^{t+1}, \lambda_C^{t+1})$, s.t. the new maximum regret r^{t+1} will be less or equal to the previous $r_s < r^t$. This holds for any iteration t , and since the regret is bound by below by zero, Alg. 1 will converge.

Theorem 4 straightforwardly implies the following:

Corollary 1. *Let a bimatrix game $G = (R, C) \in [0, 1]^{n \times n}$. Then, our algorithm will converge to a stationary point of the game $G' = (R + \lambda_R^* \cdot C, C + \lambda_C^* \cdot R)$, with $(\lambda_R^*, \lambda_C^*)$ being the final assigned λ 's.*

Note the previous results imply that the reached strategy profile $(\mathbf{x}^*, \mathbf{y}^*)$ is a stationary point given $(\lambda_R^*, \lambda_C^*)$, and the pair $(\lambda_R^*, \lambda_C^*)$ is an optimal behaviour given $(\mathbf{x}^*, \mathbf{y}^*)$.

6 Experimental Evaluation

Here, we first examine experimentally the effects of altruistic/spiteful behaviour on the regret of the modified game. We then show the potential for learning the degree of altruistic/spiteful behaviour of potential opponents, and employ the knowledge acquired to identify the most “preferred” opponent. Moreover, we showcase the potential for the *transfer of knowledge* regarding an opponent’s λ and for *transfer learning* between different game settings.¹⁰

6.1 Approximate NE under altruism and spite

Here, we present experiments to assess the value of approximate NE under various levels of altruism, and to test Alg. 1.

In our first experiments, we will compute approximate NE under various (λ_R, λ_C) values, using PGD to solve either Problem 3, or Problem 4 for different λ_R, λ_C values. Following that, we will characterize an approximate NE considering its ε value.

To begin, for each bimatrix game, we create $21 \times 21 = 441$ modified ones; each modified game has a different combination of λ values for the two players. The 441 games were derived by combining 21 different values for λ for each player. These correspond to numbers ranging from -1 to 1 in increments of 0.1 (e.g., $[-1.0, -0.9, -0.8, \dots, 0.8, 0.9, 1.0]$). For each game, we constructed a 21×21 mesh grid that plots the ε approximation for each combination of λ values. We calculate ε after PGD converged to a strategy profile (\mathbf{x}, \mathbf{y}) for both players. For calculating the strategy profile, we solve (via PGD) Problems 4 (or Problems 6).

¹⁰ The experiments were run on an Ubuntu 22.04 computer. This computer has a Ryzen 7, 16-core CPU and 16GB of ram.

Note that we address the dependence of the optimization problem's solution on the initial point by solving each problem multiple (specifically, 20) times, each with a distinct initial strategy profile (\mathbf{x}, \mathbf{y}) . Unless otherwise specified, initial strategy profiles were chosen randomly by sampling $2n$ samples (where $n > 1$ is the number of distinct actions) from a standard normal distribution, and applying softmax in order to ensure $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i = 1$. Our final results are then computed as averages over those runs.

The bimatrix game (7) below describes the payoff matrices for the *Penalty Game (PG)* in [12]. Parameter k is usually set to a high negative value to express a high penalty. In our experiment we set $k = -100$. This game has three pure NE points, (x_i, x_i) , for $i = 1, 2, 3$. However, those equilibrium points are not equivalent: e.g., the (x_2, y_2) strategy leads to a pure NE with 2 for each player, while (x_1, y_1) rewards each player with 10.

$$R = C = \begin{pmatrix} 10 & 0 & k \\ 0 & 2 & 0 \\ k & 0 & 10 \end{pmatrix}. \quad (7)$$

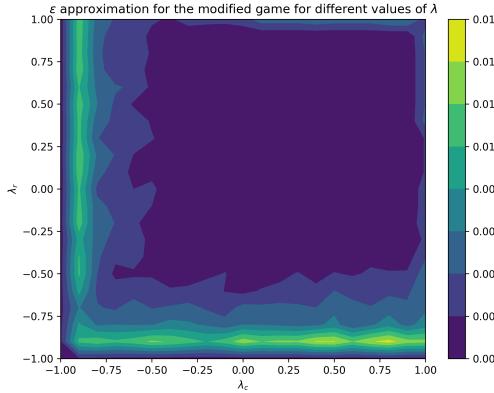


Fig. 1. The ε approximation for PG for the modified game. Results are averages over 20 runs.

Figure 1 reports the approximation of NE for different values of λ , calculated as in Definition 2, using the strategy profile after convergence. As seen in Figure 1, the addition of altruism/spite does not drastically change the already low (near-zero) approximation. In other words, PGD converges to a low approximation ε -NE for most combinations of λ values. However, when the behaviour of at least one player is described by a $\lambda \approx -0.8$, then the approximation slightly increases.

Figure 2 shows the approximations in the TS tight example [11] after the agents converged to a strategy profile solving Program 3 for all λ values combinations. For this example, we set $b = 0.3393$. We began the optimization process

for each such combination from the $\mathbf{x} = \mathbf{y} = [1 \ 0 \ 0]$ initial point, which is an approximate NE with $\varepsilon = b = 0.3393$. Here we can see that in most cases we can achieve a better approximation than the 0.3393 of the initial point in the initial game.

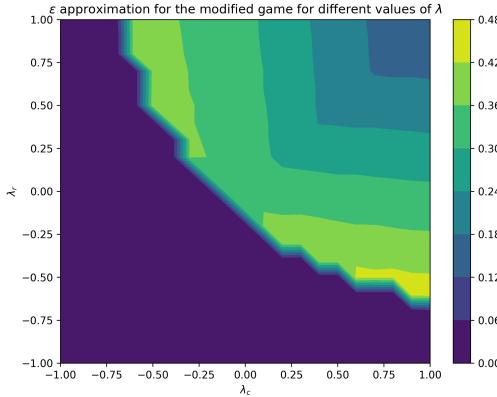


Fig. 2. Approximation for the TS tight example [11] for the modified game.

We also tested our approach with a “maximization version” of the *Prisoner’s Dilemma (PD)* game (i.e, in this version any player aims to maximizes her payoff, instead of minimizing her years in prison as in the most common PD variants). Our PD is defined as follows:

$$R = \begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}, \quad (8)$$

Fig. 3 shows the ε approximation in the modified game of (8) after converging to a strategy profile (\mathbf{x}, \mathbf{y}) (following the same simulations process described for the Penalty Game above—starting again from a random initial strategy profile (\mathbf{x}, \mathbf{y})). Notice that excellent ε -NE approximations (teal-colored region) are achieved for any combination of λ values. This implies that in the PD’s case an exact Nash equilibrium can be computed by the players for any combination of different behavioural policies (λ) of the players. This example gives insights for further research on the properties of games for which such a phenomenon holds.

Now, the following experiment tests Algorithm 1 in the first motivating example (Example of Equation 1). By running this alternating procedure, we simultaneously optimize the strategy and the “behavioural” profile of the two agents. We remind the reader that Algorithm 1 alternatively modifies the λ values and the strategies of the agents, progressively minimizing the regret ε (i.e., improving the approximation). As such, we expect ε to have a descending trajectory until it converges to a new value.

Specifically, we ran Algorithm 1 for the game presented in Equation 1, for a number of $h = 100$ iterations. The algorithm initiates with $(\lambda_R, \lambda_C) = (0, 0)$ and

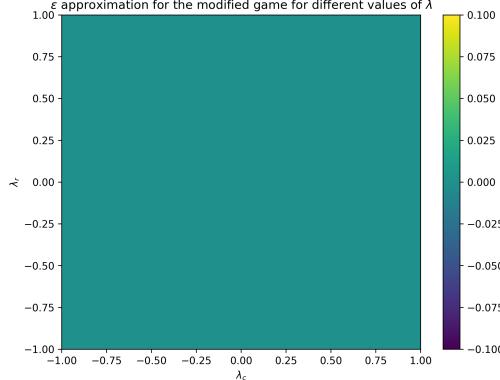


Fig. 3. The ϵ approximation for our Prisoner’s Dilemma (PD) game in its modified form. Averages over 20 runs.

strategy profile $\mathbf{x} = \mathbf{y} = [1 \ 0 \ 0]$. Figure 4 shows the change of the ϵ approximation for the modified game during a run of Algorithm 1. As observed, the approximation of the modified game indeed decreases as the algorithm fine-tunes the optimization variables appropriately. After a few iterations, it converges to a low approximation for the modified game.

Our experiments demonstrate some interesting phenomena. First, in the tight example of the TS algorithm the introduction of altruism/spite can lead players to points that have very good approximation, even $\epsilon = 0$ (purple-blue in Fig. 2). This shows that altruistic/spiteful behavior can potentially lead to more stable solutions, which is not always the case when players are selfish. Additionally, a better approximation can be achieved via Alg. 1 (Fig. 4).

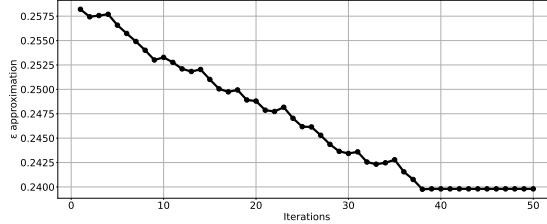


Fig. 4. Alg. 1 approximation for the example of Eq. 1.

6.2 Opponent Modeling and Opponent Selection

We now turn our attention to settings with *uncertainty* regarding the λ values of opponents. Specifically, consider a setting in which agents get the opportunity

to observe the behaviour of potential opponents over some “training” period, during which it is matched against each one of them in a *repeated* mode. This allows agents to model their opponents and infer their λ ’s. Intuitively, a learning agent is then able to “select” an opponent to face in a single-shot bimatrix game setting, given its own perception of what is beneficial to it (i.e., its own self-interest/altruism/spite point of view).

Note that, contrary to what one could simplistically assume, it is not always the case that agents would prefer to play against an altruist. Consider, e.g., the ε approximation in the TS tight example [11] (Fig. 2). An agent with $\lambda = 0$ achieves a lower approximation when playing against a more spiteful ($\lambda < 0$) opponent. Thus, specific λ -opponent choice depends on the particular game setting.

We now detail our experimental setup. We have a total of 101 agents in our setting. For simplicity, we assume only one learning agent i , with a $\lambda_i = 0$, known to all other agents j . The (100 in number) λ_j values of the j agents, are sampled from a uniform random distribution between $[-1, 1]$. Then, i plays 10^4 times against each other agent j in a (symmetric) bimatrix game whose R is randomly created from a Gaussian with $\mu = 0.5$ and $\sigma^2 = 0.2$, while $C = R^T$. In this one-shot game, each agent has 15 actions available.

Each j agent calculates at the start of the game its strategy profile \mathbf{y}^j . This strategy is calculated by solving Problem 4 given the known λ values— j knows both λ_j and λ_i . Then, in each round, it chooses an action according to its \mathbf{y}^j . At the end of each round, agents observe the selected actions.

Now, agent i has no knowledge of the λ_j of any other agent j , and intends to learn it. In principle, i can use any opponent modeling process of its choosing (e.g., a Bayesian beliefs updating one). Here it uses a simple process inspired by *fictitious play* [18]: during the course of each repeated game, it maintains an *empirical distribution* that models the frequency of j ’s actions. Let us call this distribution $\hat{\mathbf{y}}^{ij}$ (i ’s estimate of j ’s strategy \mathbf{y}^j).

Then, agent i (whose λ_i is 0) iteratively solves program 4 for all λ values between $[-1, 1]$ with ‘step’ equal to 0.1 (ex. $[-1, -0.9 \dots 0.9, 1]$). Hence, it is able to calculate a strategy $\mathbf{y}(\lambda)$ for every potential λ “behavioural type” facing $\lambda_i = 0$. Thus, i is able to compare $\hat{\mathbf{y}}^{ij}$ with every $\mathbf{y}(\lambda)$ and infer a $\hat{\lambda}_j$ value for j —i.e., sets $\hat{\lambda}_j = \operatorname{argmin}_{\lambda} \|\mathbf{y}(\lambda) - \hat{\mathbf{y}}^{ij}\|_1$. Ideally, $\hat{\lambda}_j$ is close to the actual λ_j characterizing agent j .

Having estimated a $\hat{\lambda}_j$ for every agent j , agent i can now select opponents with a specific “behavioural type” to challenge them in a new one-shot (or repeated) game. Since the bimatrix game during this experiment is random, there is no intuitive explanation on which opponents (and of which “behavioural type”) the agent i should select. Since i has $\lambda = 0$, it selects as an opponent a j whose $\hat{\lambda}_j$ value is such that i ’s reward is maximized given $\mathbf{y}(\hat{\lambda}_j)$ and $\mathbf{x}(\lambda_i)$, with $\lambda_i = 0$. In this particular game, i selects an opponent with $\hat{\lambda}_j = -0.5$, that is, an opponent perceived as rather spiteful.

Table 1 shows the reward of agent i during this experiment. In more detail, it shows the average reward achieved during the training phase, calculated across

Table 1. Average rewards during training (before opponent selection) and evaluation (after opponent selection).

Training Reward	0.5172
Evaluation Reward vs $\hat{\lambda}_j = -0.5$	0.5451
Eval. Reward vs Spiteful opponents ($\lambda_j < -0.5$)	0.5422
Eval. Reward vs Altruistic opponents ($\lambda_j > 0.5$)	0.5302

all 100 opponents, over 10^4 repeated game rounds per opponent. During training, agent i used a random strategy, while every opponent j was selecting actions according to its actual \mathbf{y} strategy, i.e. the strategy prescribed by its “behavioural type” when playing against a $\lambda = 0$ agent, with \mathbf{y} estimated via Program 4.

Moreover, Table 1 shows i ’s evaluation reward $\mathbf{x}^T R \mathbf{y}$, after selecting its j opponent according to the inferred $\hat{\lambda}_j = -0.5$ behavioural type, and where \mathbf{y} is the strategy profile of the actual λ_j opponent. We note that $\hat{\lambda}_j$ coincided with the actual λ_j of that opponent—i.e., i ’s learning process was successful. Indeed, i managed to learn the actual λ ’s of all the opponents.

For interest, Table 1 also shows evaluation reward (i) against all “s spiteful” agents with $\lambda_j < -0.5$; and (ii) against all “altruists” with $\lambda_j > 0.5$. We see that it is indeed beneficial to i to play against $\hat{\lambda}_j = -0.5$ (and to a lesser extent, against other, more spiteful agents), in this particular game.

6.3 Transfer of knowledge and transfer learning

Our agents’ ability to model their opponents and learn their altruistic/spiteful behaviour, as outlined in the experiments above, provides the potential for the *transfer of knowledge* between game environments. That is, the potential arises for an agent that has learned an opponent’s λ denoting its level of altruism (or spite) in one particular game setting, to *exploit* this knowledge in a *distinct* game setting, potentially increasing its reward in that second game compared to what it would have been had it been ignorant of the actual opponent’s λ (and thus had effectively assumed it to be equal to zero as the SIH requires). For example, consider the game

$$G = \begin{pmatrix} 4, 2 & 0, 3 \\ 3, 0 & 1, 1 \end{pmatrix}.$$

We can see that the only NE is the strategy profile with payoffs (1, 1). Let us now assume that the Row player has $\lambda_R = 0$ and *has learned* (possibly via training in a completely different game in the past) the $\lambda_C = 1$ of the Column player. Then, Row realizes that the actual game that will be played against Column, is

$$G' = \begin{pmatrix} 4, 6 & 0, 3 \\ 3, 3 & 1, 2 \end{pmatrix}.$$

We can see that the unique NE in game G' is the strategy profile with payoffs (4, 6). Thus, since Row knows the λ_C determining the respective strategy of the

Column player, she will adapt her own strategy and achieve a better payoff (i.e., a 4 instead of a 1).

Moreover, we point out that the opponent modeling process of Section 6.2 can be used for *transfer learning* purposes, in the spirit of the namesake machine learning concept [25]. That is, in our problem domain, an agent can leverage its experience from learning, at least “partially” or with some degree of confidence, the λ of an opponent in a game (e.g., via the process presented in Section 6.2), in order to accelerate its learning of that λ value in a second, distinct to the first one, game setting (via the same learning process, or, in principle, via any learning process of its choosing). For instance, transfer learning could be particularly useful in situations when the first game is a small one, while the second game is a significantly larger one (in terms of the action set and thus potentially the equilibrium strategy support set); i.e., one that would require a large number of iterations (rounds) in order to learn it.

To show this, we conducted a small experiment using two random symmetric games: a 3×3 game, denoted as A , and a 10×10 game, denoted as B . The payoff matrices for both games were sampled from a normal distribution with $\mu = 0.5$ and $\sigma = 0.2$. Agent i interacted with agent j during game A for a number of $X = 3$ times, tracking the empirical distribution \mathbf{y}_A^{ij} (i 's estimate of j 's \mathbf{y}^j strategy in A). Then, i inferred the type $\hat{\lambda}_j$ of agent j using \mathbf{y}_A^{ij} , in the same way as in Section 6.2. Consequently, using $\hat{\lambda}_j$, agent i calculated a “prior” strategy profile for j in game B via solving Problem 4. Agent i then updated that strategy profile using $Y = 53$ interactions in game B until convergence—as such, finally inferring j 's type. Now, another agent k inferred j 's type via interacting Z times *in game B only*. The results were as follows: Both agent i and k inferred the correct λ value for agent j . However, agent k did so via interacting with j for $Z = 706$ times in total, in game B only; while i was able to do so via interacting with j for a total of $X + Y = 56$ times only, during both games. This verifies the potential of transfer learning in our framework: Agent i required *an order of magnitude fewer interactions in game B compared to k*, since, unlike agent k , it had obtained an informed prior about j via interacting with it in A .

7 Related Work

As mentioned in the introduction, there exist several models in economics that define various functional forms of preferences and utility, aiming to depart from the self-interested hypothesis (SIH) [21, 7, 29]. As a result several manifestations of spite and/or altruism emerge in these models. In our work, we chose to build upon the rather intuitive definition of altruism appearing in [23]. That work provides a refinement of the NE which incorporates a small degree of reciprocal altruism, so as to encourage “friendly” behaviour and fend-off instability against small perturbations in the NE payoffs—similarly to a trembling-hand equilibrium [28], but with an altruistic twist. In our work, we are inspired by their definition of altruism to define spite in bimatrix games; and then tackle these

notions from an algorithmic game theoretic point of view, while also taking steps towards a (machine) learning treatment of this topic.

There are many game-theoretic works that consider altruism either in cooperative [20] or in non-cooperative settings [27, 4, 2]. The latter, however, either study specific game properties or player behaviour, such as specific forms of equilibria or pure NE in specific classes of games, e.g., congestion games or graph-restricted settings; but do not focus on bimatrix games nor study approximate NE as we do in our work. In most cases, existing work focuses on price of anarchy, price of stability, and (in-)efficiency of equilibria or mechanism design-relating questions [27, 8, 6, 9]. Apt and Schäfer [3], for instance, define and study the “selfishness level of strategic games”. In essence, they measure the smallest adjustment needed to be performed to player payoffs so that the socially optimal outcome is also a stable Nash equilibrium. Their work does not question the SIH per se, but effectively determines how much players would need to value the collective good so that a socially desirable outcome can be achieved via their self-interested choices. We have not so far studied such questions, but in principle they could be incorporated in our model—for instance, it could be interesting to study the form and degree of incentives that should be provided to individuals in order for them to alter their degree of altruism/spite.

8 Conclusions and future work

In our work in this paper, we move away from the SIH to provide both a theoretical and an experimental treatment of altruism and spite in the fundamental class of bimatrix games. We provide theoretical results regarding (approximate) NE under altruism/spite, showing that some better than the state-of-the-art approximations can be computed in poly-time when altruism or spite are at play; and contribute a PGD-based algorithm that optimizes towards a regret-minimizing strategies/behaviours profile. Furthermore, we explored the potential of opponent modeling and learning given uncertainty over the levels of altruism.

In future work, we intend to tackle *n-player normal form*, or *polymatrix games* [15]. Moreover, it would be interesting to add stochasticity in the game outcomes, requiring the learning of the reward matrices alongside the λ s, in a reinforcement learning fashion. Furthermore, we intend to explore the effects that *non-stationary* λ values have to convergence to (approximate) NE. More generally, it would be interesting to study the convergence of evolutionary learning/replicator dynamics in this context—e.g., identifying the behaviour dynamics’ steady states that emerge after repeated play in large, learning populations [31, 18]. As such, our work can be viewed as a stepping stone towards integrating AGT solutions with ML research—including work towards revealing the implications of bounded rationality in agent decision making.

Acknowledgments. The research described in this paper was carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union - NextGenerationEU (Implementation Body: HFRI. Project name: DEEP-REBAYES. HFRI Project Number 15430).

References

1. Adler, I.: The equivalence of linear programs and zero-sum games. *Int. J. Game Theory* **42**(1), 165–177 (2013)
2. Anagnostopoulos, A., Becchetti, L., de Keijzer, B., Schäfer, G.: Inefficiency of games with social context. In: Vöcking, B. (ed.) 6th International Symposium (SAGT 2013). vol. 8146, pp. 219–230 (2013)
3. Apt, K.R., Schäfer, G.: Selfishness level of strategic games. *Journal of Artificial Intelligence Research* **49**(1), 207–240 (2014)
4. Ashlagi, I., Krysta, P., Tennenholz, M.: Social context games. In: Internet and Network Economics, 4th International Workshop (WINE 2008). vol. 5385, pp. 675–683 (2008)
5. Bertsekas, D.P.: Nonlinear programming. *Journal of the Operational Research Society* **48**(3), 334–334 (1997)
6. Biló, V., Celi, A., Flammini, M., Gallotti, V.: Social context congestion games. *Theoretical Computer Science* **514**, 21–35 (2013), graph Algorithms and Applications: in Honor of Professor Giorgio Ausiello
7. Bolton, G.E., Ockenfels, A.: Erc: A theory of equity, reciprocity, and competition. *The American Economic Review* **90**(1), 166–193 (2000)
8. Brokkelkamp, R., Hoeijmakers, S., Schäfer, G.: Greater flexibility in mechanism design through altruism. In: International Symposium on Algorithmic Game Theory. pp. 41–59. Springer (2022)
9. Chen, P., de Keijzer, B., Kempe, D., Schäfer, G.: Altruism and its impact on the price of anarchy. *ACM Trans. Economics and Comput.* **2**(4), 17:1–17:45 (2014)
10. Chen, X., Deng, X., Teng, S.H.: Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM* **56**(3), 14:1–14:57 (2009)
11. Chen, Z., Deng, X., Huang, W., Li, H., Li, Y.: On tightness of Tsaknakis-Spirakis descent methods for approximate Nash equilibria. *Information and Computation* **293**, 105046 (2023)
12. Claus, C., Boutilier, C.: The dynamics of reinforcement learning in cooperative multiagent systems. In: Proceedings of the Fifteenth National/Tenth Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence. p. 746–752 (1998)
13. Daskalakis, C., Goldberg, P.W., Papadimitriou, C.H.: The complexity of computing a Nash equilibrium. *SIAM Journal on Computing* **39**(1), 195–259 (2009)
14. Deligkas, A., Fasoulakis, M., Markakis, E.: A polynomial-time algorithm for 1/3-approximate Nash equilibria in bimatrix games. *ACM Trans. Algorithms* **19**(4), 31:1–31:17 (2023)
15. Deligkas, A., Fearnley, J., Savani, R., Spirakis, P.G.: Computing approximate Nash equilibria in polymatrix games. *Algorithmica* **77**(2), 487–514 (2017)
16. Fasoulakis, M.: Computing approximate Nash equilibria. Ph.D. thesis, University of Warwick (2017)
17. Fehr, E., Schmidt, K.M.: The economics of fairness, reciprocity and altruism – experimental evidence and new theories. In: Handbook of the Economics of Giving, Altruism and Reciprocity, vol. 1, pp. 615–691. Elsevier (2006)
18. Fudenberg, D., Levine, D.: The Theory of Learning in Games. MIT Press (1998)
19. Hwang, S., Rey-Bellet, L.: Strategic decompositions of normal form games: Zero-sum games and potential games. *Games Econ. Behav.* **122**, 370–390 (2020)
20. Kerkemann, A.M., Cramer, S., Rothe, J.: Altruism in coalition formation games. *Annals of Mathematics and Artificial Intelligence* **92**(3), 601–629 (2024)

21. Levine, D.K.: Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics* **1**(3), 593–622 (1998)
22. Mangasarian, O.L., Stone, H.: Two-person nonzero-sum games and quadratic programming. *Journal of Mathematical Analysis and Applications* **9**(3), 348–355 (1964)
23. Marco, G.D., Morgan, J.: Slightly Altruistic Equilibria. *Journal of Optimization Theory and Applications* **137**(2), 347–362 (2008)
24. Palfrey, T., Prisbrey, J.E.: Anomalous behavior in public goods experiments: How much and why? *American Economic Review* **87**(5), 829–46 (1997)
25. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* **22**(10), 1345–1359 (2010)
26. Roth, A.E., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S.: Bargaining and market behavior in jerusalem, ljubljana, pittsburgh, and tokyo: An experimental study. *The American Economic Review* **81**(5), 1068–1095 (1991)
27. Rothe, J.: Thou shalt love thy neighbor as thyself when thou playest: Altruism in game theory. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(17), 15070–15077 (2021)
28. Selten, R.: Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory* **35**, 25–55 (1975)
29. Sobel, J.: Interdependent preferences and reciprocity. *Journal of Economic Literature* **43**(2), 392–436 (2005)
30. Tsaknakis, H., Spirakis, P.G.: An optimization approach for approximate Nash equilibria. *Internet Mathematics* **5**(4), 365–382 (2008)
31. Tuyls, K., Stone, P.: Multiagent learning paradigms. In: *Multi-Agent Systems and Agreement Technologies*, vol. 10767, pp. 3–21 (2018)