An Evolutionary Game Theoretic Analysis of Difference Evaluation Functions

Mitchell Colby
Oregon State University
colbym@engr.orst.edu

Kagan Tumer
Oregon State University
kagan.tumer@oregonstate.edu

ABSTRACT

One of the key difficulties in cooperative coevolutionary algorithms is solving the credit assignment problem. Given the performance of a team of agents, it is difficult to determine the effectiveness of each agent in the system. One solution to solving the credit assignment problem is the difference evaluation function, which has produced excellent results in many multiagent coordination domains, and exhibits the desirable theoretical properties of alignment and sensitivity. However, to date, there has been no prescriptive theoretical analysis deriving conditions under which difference evaluations improve the probability of selecting optimal actions. In this paper, we derive such conditions. Further, we prove that difference evaluations do not alter the Nash equilibria locations or the relative ordering of fitness values for each action, meaning that difference evaluations do not typically harm converged system performance in cases where the conditions are not met. We then demonstrate the theoretical findings using an empirical basins of attraction analysis.

Categories and Subject Descriptors

I.2.11 [**Computing Methodologies**]: Artificial Intelligence —Distributed Artificial Intelligence —*Multiagent systems*

1. INTRODUCTION

Coordinating multiple agents to achieve a system objective is a key step in addressing many real world problems including distributed sensor network and distributed mobile robot control [4]. One approach to developing policies for multiagent systems is the use of Cooperative Coevolutionary Algorithms (CCEAs), where multiple populations evolve in parallel, with each population evolving a policy for a particular agent or set of agents. One of the key difficulties of CCEAs is the credit assignment problem: given the performance of a team of agents, how does one assign fitness values reflecting each agent's individual performance?

GECCO '15, July 11 - 15, 2015, Madrid, Spain

 ISBN 978-1-4503-3472-3/15/07…$15.00

DOI: http://dx.doi.org/10.1145/2739480.2754770

In a CCEA, fitness assignment is context dependent and subjective, because the fitness of an agent is not only a function of the agent's policy, but also the selection and policies of its collaborating agents. In multiagent settings, this leads to the promotion of agents which perform well with a wide variety of collaborators, rather than agents which perform optimally with a particular set of collaborators. Thus, the credit assignment problem changes the problem subtly, and often leads to robust, rather than optimal, solutions [8].

In order to promote good system performance while using CCEAs, the credit assignment problem is addressed through fitness function shaping. One approach to assigning fitness is the difference evaluation function, which approximates each agent's individual contribution to the overall system performance [1]. Difference evaluation functions have provided tangible benefits in a variety of domains [1, 4].

However, to date, there has been no prescriptive theoretical analysis deriving conditions under which difference evaluations are expected to improve the probability of selecting optimal actions, limiting their applicability. This paper conducts a theoretical analysis of difference evaluations as fitness assignment operators in CCEAs, using an evolutionary game theoretic framework. The specific contributions of this paper are to:

1. Prove difference evaluations improve the probability of selecting best response actions when the expected payoffs of those actions are relatively low (Section 5.1).
2. Prove that difference evaluations do not alter the Nash equilibria locations in the system (Section 5.2).
3. Prove that difference evaluations do not alter the ordering of expected payoffs for each action (Section 5.2).
4. Provide an empirical basins of attraction analysis which showcases the results of the theorems (Section 7).

Although there have been multiple studies involving the performance benefits of difference evaluation functions, this paper provides the first prescriptive theoretical analysis of difference evaluation functions, proving conditions under which they improve the probabilities of each agent to select its best response action. The rest of this paper is organized as follows. Section 2 provides details on background information and related work, Section 3 demonstrates how difference evaluations are incorporated into an evolutionary game theoretic model, Section 4 derives normalized fitness values for the evolutionary game theoretic model, Section 5 contains the proofs of our theorems, Section 6 provides an empirical demonstration of the implications of the theorems, Section 7 provides an empirical basins of attraction analysis of the proved theorems, and Section 8 concludes the paper.

2. BACKGROUND

The following sections introduce cooperative coevolutionary algorithms, difference evaluation functions, and the evolutionary game theoretic model for cooperative coevolution.

2.1 Cooperative Coevolutionary Algorithms

Evolutionary algorithms (EAs) are a class of stochastic search algorithms, which have been shown to work well in complex domains where gradient information is not readily available. EAs typically contain two basic mechanisms: mutation and selection [12]. These mechanisms act on an initial set of candidate solutions in order to generate new solutions and to retain solutions which show improvement as evolutionary time progresses. Coevolutionary algorithms are an extension of EAs in which multiple populations evolve in parallel, and are well suited for multiagent domains [5, 15]. In a coevolutionary algorithm, the fitness of an agent depends on the interactions it has with other agents. Thus, assessing the fitness of each agent is context-sensitive and subjective [8]. We focus on *Cooperative* Coevolutionary Algorithms (CCEAs), where a group of agents succeed or fail as a team.

One key issue with CCEAs is that they tend to favor stable, rather than optimal, solutions [2, 8]. As agents are paired with multiple different sets of collaborators in CCEAs during different evolutionary time steps, agents which learn to coordinate with a wide variety of teammates tend to receive higher fitness values than agents which learn to coordinate optimally with a specific set of collaborators [8]. For example, if an optimal agent only performs well with a specific set of collaborators, then that agent will likely receive a low fitness value during evolution, and be removed from the population. The difference evaluation function aims to address the subjective nature of fitness assignments associated with CCEAs, by determining agent-specific fitness values based on the value of a particular agent's policy.

2.2 Difference Evaluation Functions

The agent-specific difference evaluation function $D_i(z)$ is defined as [1]:

$$D_i(z) = G(z) - G(z_{-i} + c_i) \tag{1}$$

where $G(z)$ is the global evaluation function, and $G(z_{-i}+c_i)$ is the global evaluation function without the effects of agent i. The term c_i is the *counterfactual*, which is used to replace agent i, and must not depend on the actions of agent i. In general, the counterfactual term is chosen to either remove the agent from the system or to replace the agent with an "average" agent. We demonstrate how a counterfactual may be chosen in Section 3. Intuitively, the difference evaluation function gives the impact of agent i on the global evaluation function, because the second term removes the portions of the global evaluation function not dependent on agent i.

Difference evaluations have two key theoretical properties which allow for improved system performance. First, any action which increases $D_i(z)$ also increases $G(z)$. This property is termed *alignment* [1]. Second, as the second term in the difference evaluation removes the effects of all agents other than agent i, the difference evaluation provides a feedback signal with much less noise than $G(z)$. This property is termed *sensitivity* [1].

Although difference evaluations have produced excellent empirical results and exhibit the desirable theoretical properties of alignment and sensitivity, the theoretical advantages of difference evaluations have not been well developed. To date, there has been no prescriptive theoretical analysis which determines under what conditions difference evaluations are expected to improve system performance. This paper addresses these theoretical questions by analyzing difference evaluations using an evolutionary game theoretic framework.

2.3 EGT Model for Cooperative Coevolution

This section introduces the Evolutionary Game Theoretic (EGT) model we use for our analysis. EGT is well suited as a model of cooperative coevolutionary algorithms, and thus provides a theoretical framework in which we can analyze the effects of different fitness assignment operators [6, 9, 10, 11, 13, 14]. This analysis is restricted to cooperative coevolutionary algorithms with two agents learning in stateless domains, and each agent having a finite number of actions. The model assumes that the populations are infinite, and that the proportions of individuals in the populations are computed at each time step during evolution. If the first agent has a finite number of n distinct actions it can take, then its population at each generation is an element of the unit simplex $\Delta^n = \{x \in [0,1]^n | \sum_{i=1}^n x_i = 1\}$. A higher value x_i corresponds to a higher probability that the agent selects action i. If the second agent has m actions to choose from, then its population at each generation is an element of the unit simplex $\Delta^m = \{y \in [0,1]^m | \sum_{j=1}^n y_j = 1\}$. The EGT model for CCEAs we use is the discrete time replicator dynamics with fitness proportional selection, defined as in [8, 13]:

$$u_i^{(t),c} = \sum_{j=1}^{m} c_{ij} y_j^{(t)} \tag{2}$$

$$w_j^{(t),c} = \sum_{i=1}^{n} c_{ij} x_i^{(t)} \tag{3}$$

$$x_i^{(t+1),c} = \left(\frac{u_i^{(t),c}}{\sum_{k=1}^n x_k^{(t)} u_k^{(t),c}} \right) x_i^{(t)} \tag{4}$$

$$y_j^{(t+1),c} = \left(\frac{w_j^{(t),c}}{\sum_{k=1}^m y_k^{(t)} w_k^{(t),c}} \right) y_j^{(t)} \tag{5}$$

Where:

- $u_i^{(t),c}$ is the fitness of agent 1 taking action i at time t, while using the global payoff matrix.
- $w_j^{(t),c}$ is the fitness of agent 2 taking action j at time t, while using the global payoff matrix.
- $x_i^{(t)}$ is the probability agent 1 takes action i at time t.
- $y_j^{(t)}$ is the probability agent 2 takes action j at time t.
- $x_i^{(t+1),c}$ is the updated probability agent 1 takes action i at time $t+1$, when the update is completed using the global payoff matrix.
- $y_j^{(t+1),c}$ is the updated probability agent 2 takes action j at time $t+1$, when the update is completed using the global payoff matrix.

Note that the fitness values of each action are simply the expected payoffs of taking each action, given the collaborating agent's probability distribution for action selection. For this analysis, we make the following assumptions:

A1) All elements of the payoff matrix are non-negative. If there are negative terms in a payoff matrix, a positive constant is added to each element to ensure each element is non-negative.

A2) There are at least two distinct elements in the payoff matrix (not all elements have the same value).

Assumption A1 is a common assumption in evolutionary game theory; non-negative payoff matrix elements ensure that the system remains invariant in the simplex [7, 14]. Assumption A2 is needed in the proofs, but it is worth noting that if this assumption does not hold, then we have a trivial payoff matrix where every element has equal value.

3. DIFFERENCE PAYOFF MATRICES

We now demonstrate how difference evaluation functions are incorporated into the EGT model. Suppose that each agent is given agent-specific feedback rather than global feedback. We can directly apply the difference evaluation function from Equation 1 to the global payoff matrix C, creating the agent-specific *difference payoff matrices* D^1 and D^2, which are defined as:

$$d^1_{ij} = c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max} \tag{6}$$

$$d^2_{ij} = c_{ij} - \frac{1}{m}\sum_{k=1}^{m} c_{ik} + c_{max} \tag{7}$$

Note that to compute the counterfactual term, we calculate the average payoff for an agent across all of its potential actions, given the action of the collaborating agent. In general, it is desirable to choose a counterfactual which effectively removes an agent from the system. However, in a game-theoretic setting, agents cannot be removed, because each agent must select an action to find a joint payoff. So, rather than removing the agent from the system, the counterfactual term in the difference payoff matrices replace the agent with an average agent with a uniform random policy. The term c_{max} is added to ensure that all elements of the difference payoff matrix are non-negative, where:

$$c_{max} = max_{ij}\left\{|c_{ij}|\right\}$$

The fitness assignment operators from the EGT model (Equations 2 and 3) are altered by directly incorporating the difference payoff matrices from Equations 6 and 7. The EGT model for CCEAs with difference evaluations is thus:

$$u_i^{(t),d} = \sum_{j=1}^{m} d^1_{ij} y_j^{(t)} \tag{8}$$

$$w_j^{(t),d} = \sum_{i=1}^{n} d^2_{ij} x_i^{(t)} \tag{9}$$

$$x_i^{(t+1),d} = \left(\frac{u_i^{(t),d}}{\sum_{k=1}^{n} x_k^{(t)} u_k^{(t),d}}\right) x_i^{(t)} \tag{10}$$

$$y_j^{(t+1),d} = \left(\frac{w_j^{(t),d}}{\sum_{k=1}^{m} y_k^{(t)} w_k^{(t),d}}\right) y_j^{(t)} \tag{11}$$

Where:

- $u_i^{(t),d}$ is the fitness of agent 1 taking action i at time t, while using the difference payoff matrix.
- $w_j^{(t),d}$ is the fitness of agent 2 taking action j at time t, while using the difference payoff matrix.
- $x_i^{(t+1),d}$ is the updated probability agent 1 takes action i at time $t+1$, when the update is completed using the difference payoff matrix.
- $y_j^{(t+1),d}$ is the updated probability agent 2 takes action j at time $t+1$, when the update is completed using the difference payoff matrix.

Note that the only difference between the traditional EGT model and the EGT model incorporating difference evaluations occurs in the fitness assignment stage. Rather than using the global payoff matrix to assign fitness, we use the difference payoff matrices.

4. NORMALIZED FITNESS VALUES

As the difference evaluation function alters the elements of the payoff matrix, we must normalize fitness values in order to compare performance when using difference payoff matrices versus the global payoff matrix. In the following sections we derive normalized fitness values for agents using both global and difference evaluation functions.

4.1 Global Evaluation Function

Fitness values for each agent using the global payoff matrix are defined in Equations 2 and 3. We normalize these fitness values such that $\sum_{i=1}^{n} \bar{u}_i^{(t),c} = 1$ and $\sum_{j=1}^{m} \bar{w}_j^{(t),c} = 1$. The normalized expected payoff for the first agent taking action i and using the global evaluation function at time t is:

$$\bar{u}_i^{(t),c} = \frac{u_i^{(t),c}}{\sum_{k=1}^{n} u_k^{(t),c}} \tag{12}$$

$$= \frac{\sum_{j=1}^{m} c_{ij} y_j^{(t)}}{\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)}} \tag{13}$$

A similar derivation for the second agent yields:

$$\bar{w}_j^{(t),c} = \frac{\sum_{i=1}^{n} c_{ij} x_i^{(t)}}{\sum_{i=1}^{n}\sum_{k=1}^{m} c_{ik} x_i^{(t)}} \tag{14}$$

4.2 Difference Evaluation Function

We now define the fitness for an agent using the difference payoff matrices in terms of the global payoff matrix. The fitness for the first agent taking the action i while using the difference evaluation function at time t is:

$$u_i^{(t),d} = \sum_{j=1}^{m} d^1_{ij} y_j^{(t)} \tag{15}$$

$$= \sum_{j=1}^{m}\left(c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max}\right) y_j^{(t)} \tag{16}$$

$$= u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max} \tag{17}$$

The normalized fitness for the first agent taking action i

while using the difference evaluation at time t is:

$$\bar{u}_i^{(t),d} = \frac{u_i^{(t),d}}{\sum_{k=1}^{n} u_i^{(t),d}} \tag{18}$$

$$= \frac{u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max}}{\sum_{l=1}^{n}\left(u_l^{(t),c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max}\right)} \tag{19}$$

$$= \frac{u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max}}{\sum_{l=1}^{n} u_l^{(t),c} - \sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + n \cdot c_{max}} \tag{20}$$

$$= \frac{u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n} u_k^{(t),c} + c_{max}}{\sum_{l=1}^{n} u_l^{(t),c} - \sum_{k=1}^{n} u_k^{(t),c} + n \cdot c_{max}} \tag{21}$$

$$= \frac{u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n} u_k^{(t),c} + c_{max}}{n \cdot c_{max}} \tag{22}$$

A similar derivation yields the normalized fitness for the second agent taking action j and using the difference evaluation at time t:

$$\bar{w}_j^{(t),d} = \frac{w_j^{(t),c} - \frac{1}{m}\sum_{k=1}^{m} w_k^{(t),c} + c_{max}}{m \cdot c_{max}} \tag{23}$$

5. DIFFERENCE EVALUATIONS THEORY

In the following sections, we derive the conditions under which difference evaluations improve the probability of selecting best response actions. We then prove that in cases where these conditions are not met, difference evaluations do not negatively affect the game.

5.1 Optimal Payoff Theory

We define the joint expected system payoff at time t as:

$$E_{tot}^{(t)}[C] = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} x_i^{(t)} y_j^{(t)} \tag{24}$$

We now prove that in cases where the optimal action (corresponding to the optimal Nash equilibrium) has a relatively low expected payoff, difference evaluations result in higher probabilities of selecting best response actions compared to the global evaluation.

THEOREM 1. *If the fitness values for the best response actions i^* and j^* are less than the joint expected system payoff, then difference evaluations result in higher probabilities of selecting the best response actions as compared to the global evaluation function.*

$$E_{tot}^{(t)}[C] > u_{i^*}^{(t),c} \Rightarrow x_{i^*}^{(t+1),d} > x_{i^*}^{(t+1),c} \tag{25}$$

PROOF. Starting with Equation 25, we have that:

$$E_{tot}^{(t)}[C] > u_{i^*}^{(t),c}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} x_i^{(t)} y_j^{(t)} > u_{i^*}^{(t),c}$$

Noting that $u_i^{(t),c} = \sum_{j=1}^{m} c_{ij} y_j^{(t)}$, we have that:

$$\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}$$

We now multiply both sides of the inequality by a positive constant A:

$$A\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > A u_{i^*}^{(t),c}$$

Noting that $\sum_{i=1}^{n} A x_i^{(t)} = A$, we have that:

$$A\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\sum_{i=1}^{n} A x_i^{(t)}$$

We add $u_{i^*}^{(t),c}\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} - u_{i^*}^{(t),c}\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} = 0$ to the right hand side of the inequality in order to allow factoring of terms, yielding:

$$A\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\left[\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} + \sum_{i=1}^{n} A x_i^{(t)} - \sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)}\right]$$

$$\Rightarrow A\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\left[\sum_{i=1}^{n}\left(u_i^{(t),c} + A\right) x_i^{(t)} - \sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)}\right]$$

$$\Rightarrow A\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} + u_{i^*}^{(t),c}\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\sum_{i=1}^{n}\left(u_i^{(t),c} + A\right) x_i^{(t)}$$

$$\Rightarrow \left(u_{i^*}^{(t),c} + A\right)\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\sum_{i=1}^{n}\left(u_i^{(t),c} + A\right) x_i^{(t)} \tag{26}$$

We now focus on the term A from Equation 26. Recall that:

$$u_i^{(t),d} = \sum_{j=1}^{m}\left(c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max}\right) y_j^{(t)}$$

$$= \sum_{j=1}^{m} c_{ij} y_j^{(t)} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + \sum_{j=1}^{m} c_{max} y_j^{(t)}$$

$$= u_i^{(t),c} - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max}$$

We thus define A as:

$$A = -\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j^{(t)} + c_{max}$$

Note that A is strictly positive by assumptions A1 and A2 from Section 2.3. With this definition of A, we have that:

$$u_i^{(d),t} = u_i^{(c),t} + A \tag{27}$$

Combining Equations 26 and 27 yields:

$$u_{i^*}^{(t),d}\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)} > u_{i^*}^{(t),c}\sum_{i=1}^{n} u_i^{(t),d} x_i^{(t)}$$

$$\Rightarrow \frac{u_{i^*}^{(t),d}}{\sum_{i=1}^{n} u_i^{(t),d} x_i^{(t)}} > \frac{u_{i^*}^{(t),c}}{\sum_{i=1}^{n} u_i^{(t),c} x_i^{(t)}} \tag{28}$$

Note that the terms in the inequality from Equation 28 are equivalent to the coefficients in the population update rules from Equations 4 and 10. We thus have:

$$x_{i^*}^{(t+1),d} > x_{i^*}^{(t+1),c} \tag{29}$$

A similar derivation for the second agent yields:

$$E_{tot}^{(t)}[C] > w_{j^*}^{(t),c} \Rightarrow y_{j^*}^{(t+1),d} > y_{j^*}^{(t+1),c} \tag{30}$$

□

Thus, if the fitness values for the optimal actions i^* and j^* are less than the joint expected system payoff (i.e. the expected payoff of the optimal actions is relatively low), then difference evaluations result in higher probabilities of selecting the optimal actions than the global evaluation function does. Equations 25 and 30 prescribe conditions under which difference evaluations will result in better system performance than the global evaluation function. For clarity, Section 6 provides a numerical example demonstrating the results of this theorem.

5.2 Game Characteristics Theory

In Section 5.1, we proved that in cases where the expected payoffs of the optimal actions are relatively low, difference evaluations increase the probability of selecting the optimal actions compared to global evaluations. However, this also means that agents using difference evaluations have lower probabilities of selecting optimal actions in cases where the expected payoff of the optimal actions are relatively high. Even though this is the case, difference evaluations typically do not harm converged system performance when the prescribed condition from Theorem 1 does not hold. We now provide two theorems which give the intuition for why this is the case.

We first prove that all Nash equilibria in the game when using the global payoff matrix remain Nash equilibria when using difference payoff matrices. This theorem demonstrates that attractor points in the population space are not altered when applying difference evaluation functions.

THEOREM 2. *All Nash equilibria remain Nash equilibria when agents use difference payoff matrices for feedback.*

PROOF. A Nash equilibrium (i^*, j^*) in a symmetric game with a payoff matrix C is defined as:

$$c_{i^*j^*} > c_{ij^*} \ \forall \ i \neq i^* \tag{31}$$

$$c_{i^*j^*} > c_{i^*j} \ \forall \ j \neq j^* \tag{32}$$

When agents are using difference evaluations, the game is not symmetric, so a Nash equilibrium (i^*, j^*) is defined as:

$$d^1_{i^*j^*} > d^1_{ij^*} \ \forall \ i \neq i^* \tag{33}$$

$$d^2_{i^*j^*} > d^2_{i^*j} \ \forall \ j \neq j^* \tag{34}$$

Suppose we have a Nash equilibrium (i^*, j^*) that satisfies Equations 31 and 32. We now demonstrate that this implies that Equations 33 and 34 are also satisfied. We begin with Equation 31, and add $-\frac{1}{n}\sum_{k=1}^{n} c_{kj^*} + c_{max}$ to both sides of the inequality:

$$c_{i^*j^*} > c_{ij^*} \ \forall \ i \neq i^*$$

$$c_{i^*j^*} - \frac{1}{n}\sum_{k=1}^{n} c_{kj^*} + c_{max} > c_{ij^*} - \frac{1}{n}\sum_{k=1}^{n} c_{kj^*} + c_{max} \ \forall i \neq i^*$$

Noting that $d^1_{ij} = c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max}$, we have that:

$$d^1_{i^*j^*} > d^1_{ij^*} \ \forall i \neq i^*$$

We can similarly start from Equation 32 and derive that:

$$d^2_{i^*j^*} > d^2_{i^*j} \ \forall j \neq j^*$$

□

So, difference evaluations do not alter the location of Nash equilibria from the original game using global evaluations. Thus, difference evaluations do not alter the location of attractor points in the population space.

We now prove that the relative ordering of fitness values is identical whether difference or global evaluations are used.

THEOREM 3. *If:*

$$u_a^{(t),c} > u_{a'}^{(t),c} \tag{35}$$

Then:

$$u_a^{(t),d} > u_{a'}^{(t),d}$$

PROOF. We start with Equation 35:

$$u_a^{(t),c} > u_{a'}^{(t),c}$$

$$\Rightarrow \sum_{j=1}^{m} c_{aj} y_j > \sum_{j=1}^{m} c_{a'j} y_j$$

Adding $-\frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j + \sum_{j=1}^{m} c_{max} y_j$ to both sides of the inequality yields:

$$\sum_{j=1}^{m} c_{aj} y_j - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j + \sum_{j=1}^{m} c_{max} y_j > \sum_{j=1}^{m} c_{a'j} y_j - \frac{1}{n}\sum_{k=1}^{n}\sum_{j=1}^{m} c_{kj} y_j + \sum_{j=1}^{m} c_{max} y_j$$

$$\Rightarrow \sum_{j=1}^{m} \left(c_{aj} - \frac{1}{n}\sum_{k=1}^{m} c_{kj} + c_{max} \right) y_j > \sum_{j=1}^{m} \left(c_{a'j} - \frac{1}{n}\sum_{k=1}^{m} c_{kj} + c_{max} \right) y_j$$

Noting that $d^1_{ij} = c_{ij} - \frac{1}{n}\sum_{k=1}^{n} c_{kj} + c_{max}$, we have that:

$$\sum_{j=1}^{m} d^1_{aj} y_j > \sum_{j=1}^{m} d^1_{a'j} y_j$$

$$\Rightarrow u_a^{(t),d} > u_{a'}^{(t),d}$$

We can conduct a similar proof for the second agent, yielding:

$$w_a^{(t),c} > w_{a'}^{(t),c} \Rightarrow w_a^{(t),d} > w_{a'}^{(t),d}$$

□

Thus, if an action a has a higher fitness than action a' when using the global evaluation, then a also has a higher fitness than a' when using the difference evaluation.

From Theorem 1, we know that in cases where the optimal action has a low expected payoff, then difference evaluations improve system performance. When the conditions from Theorem 1 are not satisfied (i.e. the optimal actions have relatively high payoffs), Theorems 2 and 3 imply that difference evaluations will not alter the location of the optimal Nash equilibrium, and that the optimal actions will still have relatively high payoffs when using difference evaluations. Thus, in cases where the conditions from Theorem 1 do not hold, then difference evaluations should not negatively impact system performance.

6. PAYOFF MATRIX ANALYSIS

In Section 5.1, we proved that difference evaluations increase the probability of selecting the optimal action if the expected payoff of the optimal action is lower than the joint expected system payoff. To illustrate the effects of the difference evaluation on the dynamics of a coordination game, we will analyze how difference evaluations impact the *penalty game*, a game with a variable parameter which can alter the

risk associated with the optimal Nash equilibria [3, 8]. The penalty game is defined by the joint payoff matrix:

$$C = \begin{pmatrix} 10 & 0 & p \\ 0 & 2 & 0 \\ p & 0 & 10 \end{pmatrix} \tag{36}$$

where $p \in (-\infty, 10)$. The penalty game has three Nash equilibria: $(1,1)$, $(2,2)$, and $(3,3)$. The $(2,2)$ Nash equilibrium is suboptimal, but may be less risky depending on the value of the penalty term p. For example, if $p = -10$, then the $(2,2)$ Nash equilibrium is more forgiving if one agent deviates from its strategy. As p becomes smaller, the risk associated with the optimal Nash equilibria increases. Conversely, as p approaches 10, the risk associated with the optimal Nash equilibrium is minimized.

We can vary the value of p to alter whether or not the prescriptive conditions from Theorem 1 hold, as low values of p result in low expected payoffs for the optimal actions. We assume the agents have just started to learn and have no knowledge about the payoff matrix. As such, we assume the agents initially select their actions by sampling from a uniform random distribution. Further, we assume that the penalty term p is defined as $p = -10$. Thus, the fitness values for each action the first agent may take are:

$$u_{a_1}^{(t),c} = \frac{10 + 0 - 10 + 30}{3} = 10.0 \tag{37}$$

$$u_{a_2}^{(t),c} = \frac{0 + 2 + 0 + 30}{3} = 10.67 \tag{38}$$

$$u_{a_3}^{(t),c} = \frac{-10 + 0 + 10 + 30}{3} = 10.0 \tag{39}$$

Note that 10 is added to each element of the payoff matrix to ensure all values are non-negative. This ensures that under fitness proportional selection, the populations will remain invariant in the unit simplex. We now find the joint expected system payoff using the fitness values (expected payoffs for each action) found above:

$$\begin{aligned} E_{tot}^{(t)} &= \sum_{i=1}^{3}\sum_{j=1}^{3} c_{ij} x_i y_j \\ &= \sum_{i=1}^{3} \frac{10.0 + 10.67 + 10.0}{3} \\ &= 10.22 \end{aligned}$$

We see that $u_{a_1}^{(t),c} < E_{tot}^{(t)}$, so from Theorem 1, we know that difference evaluations will lead to a higher probability of selecting the optimal action than global evaluations will. The optimal action has a relatively low fitness, because as $y_1 = y_3$ we have the optimal Nash equilibrium payoff and penalty term canceling each other out when computing the expected payoff of the optimal action. In cases such as this where the optimal action has a low expected payoff, we know that difference evaluations increase the probability of selecting the optimal action. The normalized fitness values for the first agent taking each action are:

$$\bar{u}_{a_1}^{(t),c} = \frac{10.0}{10.0 + 10.67 + 10.0} = 0.326 \tag{40}$$

$$\bar{u}_{a_2}^{(t),c} = \frac{10.67}{10.0 + 10.67 + 10.0} = 0.348 \tag{41}$$

$$\bar{u}_{a_3}^{(t),c} = \frac{10.0}{10.0 + 10.67 + 10.0} = 0.326 \tag{42}$$

The difference payoff matrix for the first agent playing the penalty game is found by applying Equation 6 to the penalty game payoff matrix:

$$D_1 = \begin{pmatrix} 30 & 19.33 & 10 \\ 20 & 21.33 & 20 \\ 10 & 19.33 & 30 \end{pmatrix} \tag{43}$$

Retaining the assumption that each agent plays a uniform random strategy, the fitness values for each action that may be taken by the first agent are:

$$u_{a_1}^{(t),d} = \frac{30 + 19.33 + 10}{3} = 19.78 \tag{44}$$

$$u_{a_2}^{(t),d} = \frac{20 + 21.33 + 20}{3} = 20.44 \tag{45}$$

$$u_{a_3}^{(t),d} = \frac{10 + 19.33 + 30}{3} = 19.78 \tag{46}$$

In this case, the normalized fitness values for the first agent taking each action are:

$$\bar{u}_{a_1}^{(t),d} = \frac{19.78}{19.78 + 20.44 + 19.78} = 0.330 \tag{47}$$

$$\bar{u}_{a_2}^{(t),d} = \frac{20.44}{19.78 + 20.44 + 19.78} = 0.340 \tag{48}$$

$$\bar{u}_{a_3}^{(t),d} = \frac{19.78}{19.78 + 20.44 + 19.78} = 0.330 \tag{49}$$

From Theorem 1, we know that in the above example, difference evaluations result in a higher probability of selecting optimal actions. Comparing Equations 40-42 and 47-49, we see that the fitness values for the optimal actions are higher when the agent uses difference evaluation functions. Although the difference between these values is relatively small, it is of note that these differences occur after only one iteration of the EGT model. As seen in the next Section, these differences in fitness do lead to difference evaluations resulting in better performance than global evaluations.

7. EMPIRICAL RESULTS

In this section, we analyze the basins of attraction created by difference and global evaluations in the penalty game with varying values of p. We simulate the EGT model for global evaluations using Equations 2-5, and the EGT model for difference evaluations using Equations 8-11. We run 5000 simulations, using different initial population vectors (uniformly distributed across joint population space) for each simulation. Simulations are run for either 5000 generations or until the probability of selecting an action exceeds 0.995. As fitness proportional selection requires individuals to have non-negative fitness values, we add a constant term in the payoff matrix C to ensure all payoffs are non-negative.

We use the technique developed in [9] to visualize the basins of attraction. The simplex for each population is projected onto a one dimensional line segment. The line contains six regions, where all points in a given region have the same ordering of genotype proportion (1s are more common than 2s, which are more common than 3s, for example). Combining all six of these regions provides a one dimensional representation of the unit simplex. A cartesian product of two of these simplex projections provides a two dimensional representation of a two agent population space, as shown in Figure 1. Within this two dimensional space, we can visualize the basins of attraction for agents using either difference or global evaluation functions. For clarity, we have labeled

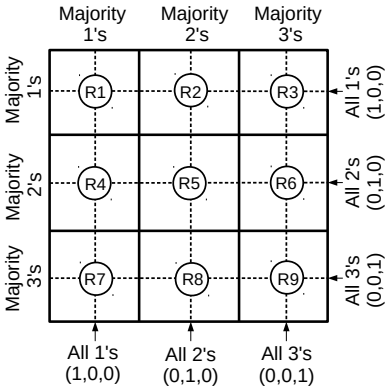

Figure 1: Visualization of the cartesian product of two simplexes, for visualing basins of attraction.

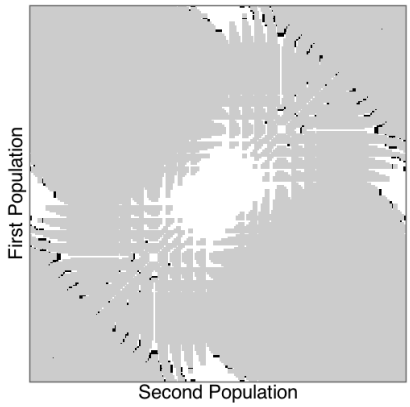

Figure 2: Basins of attraction for penalty game when $p = -50$. Gray areas: both difference and global evaluations lead to optimal Nash equilibria. Black areas: difference evaluations lead to optimal Nash equilibria, but global evaluations do not. White areas: neither difference or global evaluations lead to optimal Nash equilibra

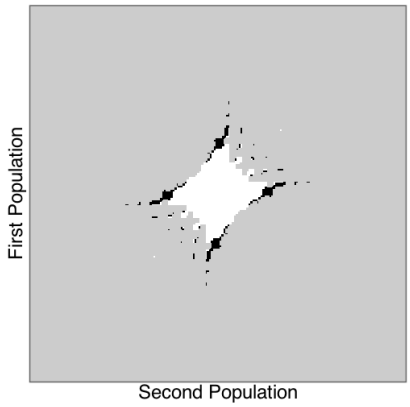

Figure 3: Basins of attraction for penalty game when $p = 0$. Gray, black, and white areas are defined as in Figure 2.

each region of the space from regions $R1$ to $R9$. See [9] for more details on the visualization process.

We analyze the basins of attraction for the penalty game when $p = -50$ and when $p = 0$. When $p = -50$, the risk associated with the optimal Nash equilibria is very high, which makes reaching the optimal Nash equilibria difficult, because both agents must consistently coordinate in order to avoid the large penalty. When $p = 0$, the optimal Nash equilibria are not difficult to reach, because there is no risk associated with the optimal actions. The values of $p = -50$ and $p = 0$ were chosen because they demonstrate three distinct cases in which difference evaluations improve system performance in the penalty game. Figure 2 shows the basins of attraction when agents use either global or difference evaluation functions and $p = -50$. Figure 3 shows the basins of attraction when agents use either global or difference evaluation functions and $p = 0$.

In Figures 2 and 3, gray areas indicate the basins of attraction around the optimal Nash equilibria for agents using either global or difference evaluation functions. Black areas indicate the regions where agents using difference evaluations converge to optimal Nash equilibria, while agents using global evaluations converge to the suboptimal Nash equilibrium. The white area indicates the basin of attraction for the suboptimal Nash equilibria $(2, 2)$ for agents using either difference or global evaluations.

As seen in Figures 2 and 3, difference evaluations improve system performance by expanding the basins of attraction around optimal Nash equilibria in three types of situations. The first case is seen in regions $R2$, $R4$, $R6$, and $R8$ of Figure 2. In this case, one agent is likely to select the suboptimal action (action 2), while the collaborating agent is likely to select an optimal action (action 1 or 3). When one agent selects the optimal action, the other agent is unlikely to do so, resulting in a poor expected payoff for the optimal action. This first case exists when $p = -50$, but not when $p = 0$. When $p = 0$, there is no penalty for an agent deviating from the optimal Nash equilibrium, meaning that the expected payoff of the optimal action is not significantly decreased.

The second case where difference evaluations are beneficial is seen in regions $R3$ and $R7$ of Figure 2. Here, one agent is likely to select an optimal action, while the collaborating agent is likely to select the mismatched optimal action. In this case, the expected payoff of the optimal action is low, because the expected payoff of the optimal action is dominated by the penalty term. As in the first case, this case only appears when $p = -50$, not when $p = 0$. This is because when $p = 0$, the penalty term for deviating from the optimal Nash equilibria is no worse than the penalty for deviating from the suboptimal Nash equilibrium, so the penalty does not push agents towards suboptimal solutions.

The third case where difference evaluations are beneficial is seen in region $R5$ of Figure 3. In this case, both agents are likely to select the suboptimal action 2. Here, the expected payoffs of the optimal actions are low, because they are dominated by the 0 terms in the payoff matrix. This improvement is not seen when $p = -50$, because the large penalty dramatically increases the risk associated with leaving the $(2, 2)$ suboptimal Nash equilibrium.

In all three cases where difference evaluations improve system performance by expanding the basins of attraction around the optimal Nash equilibria, we see that the fitness

(expected payoff) associated with the optimal action is low, which is consistent with Theorem 1. We also see that in cases where the expected payoff of the optimal action is not relatively low, both difference and global evaluation functions lead to converging to the optimal Nash equilibria, which is consistent with Theorems 2 and 3. These results demonstrate that in cases where the optimal Nash equilibrium is deceptive, difference evaluations improve the probability of selecting optimal actions, allowing agents using difference evaluations to reach the optimal Nash equilibria while agents using global evaluations converge to a suboptimal Nash equilibrium. In cases where the optimal Nash equilibria are not deceptive, agents will converge to optimal policies whether they use difference evaluations or global evaluations. We see that in some cases difference evaluations improve system performance, but they never harm system performance.

8. DISCUSSION AND CONCLUSION

Difference evaluations are effective fitness assignment operators in cooperative coevolutionary algorithms, as they provide agent-specific feedback related to the agent's impact on the overall system performance. Difference evaluations have been shown to provide superior results in a variety of multiagent coordination domains. However, to date, there has been no prescriptive theoretical analysis which derives conditions under which difference evaluations are expected to improve the probability of selecting optimal actions.

In this paper, we prove that if the fitness of the optimal action is lower than the joint expected system payoff, then difference evaluations improve the probability of selecting optimal actions. Thus, in cases where the optimal Nash equilibrium is deceptive, then difference evaluations improve system performance. Next, we proved that in cases where this condition is not met, difference evaluations do not negatively impact system performance, because they do not alter the locations of the Nash equilibria, or the relative ordering of the fitness values for each action. This means that in games where the optimal Nash equilibrium is not deceptive, then both difference and global evaluations resulting in high fitnesses for optimal actions, meaning that both fitness assignment operators promote the optimal action.

We simulated the evolutionary game theoretic model in the penalty game, and showed that the basins of attraction around optimal Nash equilibria are expanded when the conditions from Theorem 1 hold, and that difference evaluations do not negatively affect converged performance when the conditions from Theorem 1 do not hold. We find that in cases where the optimal Nash equilibria are deceptive, difference evaluations are helpful; in cases where the optimal Nash equilibria are not deceptive, difference evaluations do not harm system performance.

Future work involves extending this theory to more than two agents. The EGT framework easily extends to these cases, and early analysis shows Theorems 1-3 hold for three or more agents. We also plan on performing a theoretical analysis for multiagent reinforcement learning systems, as well as investigating different sets of replicator dynamics.

Acknowledgements

We would like to thank Dr. Karl Tuyls and Dr. Daniel Hennes for their valuable feedback on this research. This work was partially supported under NETL DE-FE0012302.

9. REFERENCES

[1] A. Agogino and K. Tumer. Analyzing and Visualizing Multiagent Rewards in Dynamic and Stochastic Environments. *Journal of Autonomous Agents and Multi-Agent Systems*, 17(2):320–338, 2008.

[2] A. Bucci and J. Pollack. Thoughts on Solution Concepts. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, GECCO '07, pages 434–439, New York, NY, USA, 2007. ACM.

[3] C. Claus and C. Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. In *In Proceedings of National Conference on Artificial Intelligence (AAAI-98*, pages 746–752, 1998.

[4] M. Colby and K. Tumer. Shaping Fitness Functions for Coevolving Cooperative Multiagent Systems. In *Proceedings of the 11th International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 425–432, Valencia, Spain, 2012.

[5] S. Ficici, O. Melnik, and J. Pollack. A Game-Theoretic and Dynamical-Systems Analysis of Selection Methods in Coevolution. *Evolutionary Computation, IEEE Transactions on*, 9(6):580 – 602, Dec 2005.

[6] D. Hennes, D. Bloembergen, M. Kaisers, K. Tuyls, and S. Parsons. Evolutionary advantage of foresight in markets. In *Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, GECCO '12, pages 943–950, New York, NY, USA, 2012. ACM.

[7] J. Hofbauer and K. Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, 1998.

[8] L. Panait, K. Tuyls, and S. Luke. Theoretical Advantages of Lenient Learners: An Evolutionary Game Theoretic Perspective. *J. Mach. Learn. Res.*, 9:423–457, June 2008.

[9] L. Panait, R. Wiegand, and S. Luke. A visual demonstration of convergence properties of cooperative coevolution. In *Parallel Problem Solving from Nature (PPSN 04)*, pages 892–901, 2004.

[10] K. Tuyls, A. Nowe, T. Lenaerts, and B. Manderick. An evolutionary game theoretic perspective on learning in multi-agent systems. *Synthese*, 139(2):pp. 297–330, 2004.

[11] K. Tuyls and S. Parsons. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7):406 – 416, 2007. Foundations of Multi-Agent Learning.

[12] D. Whitley. An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls. *Information and Software Technology*, 43(14):817 – 831, 2001.

[13] R. Wiegand. *An Analysis of Cooperative Coevolutionary Algorithms*. PhD thesis, George Mason University, 2004.

[14] R. P. Wiegand, W. C. Liles, and K. A. De Jong. Analyzing cooperative coevolution with evolutionary game theory. In *Proceedings of the Congress on Evolutionary Computation on 2002.*, CEC '02, pages 1600–1605, Washington, DC, USA, 2002. IEEE Computer Society.

[15] C. Yong and R. Miikkulainen. Cooperative Coevolution Of Multi-Agent Systems. Technical Report AI07-338, Department of Computer Sciences, The University of Texas at Austin, 2001.