A Novel Method for Automatic Strategy Acquisition in N-player Non-zero-sum Games

S. Phelps
Dept. of Computer Science
University of Liverpool
Liverpool L69 3BX
United Kingdom
sphelps@csc.liv.ac.uk

M. Marcinkiewicz
Dept. of Computer Science
Graduate Center
City University of New York
365, 5th Avenue
New York, NY 10016
USA
marek@sci.brooklyn.cuny.edu

S. Parsons
Dept. of Computer Science
City University of New York
365, 5th Avenue
New York, NY 10016, USA
parsons@sci.brooklyn.cuny.edu

ABSTRACT

We present a novel method for automatically acquiring strategies for the double auction by combining evolutionary optimization together with a principled game-theoretic analysis. Previous studies in this domain have used standard co-evolutionary algorithms, often with the goal of searching for the "best" trading strategy. However, we argue that such algorithms are often ineffective for this type of game because they fail to embody an appropriate game-theoretic *solution-concept*, and it is unclear, what, if anything, they are optimizing. In this paper, we adopt a more appropriate criterion for success from evolutionary game-theory based on the likely adoption-rate of a given strategy in a large population of traders, and accordingly we are able to demonstrate that our evolved strategy performs well.

Keywords

auctions and electronic markets, multi-agent evolution, adaptation and learning, game theoretic foundations of agent systems

1. INTRODUCTION

The automatic discovery of game-playing strategies has long been considered a central problem in Artificial Intelligence. The standard technique in evolutionary computing for discovering new strategies is *co*-evolution, in which the fitness of each individual in an evolving population of strategies is assessed relative to other individuals in that population by computing the payoffs obtained when the selected individuals interact. Co-evolution can sometimes result in *arms-races*, in which the complexity and robustness of strategies in the population increases as they counter-adapt to adaptations in their opponents.

Often, however, co-evolutionary learning can fail to converge on robust strategies. In this paper we explore some of the limitations of current co-evolutionary algorithms, and we introduce a novel method for automated strategy-acquisition by combining a game-theoretic analysis together with an evolutionary search involving a single non-coevolving population. The novel aspect of our method is that rather than using payoffs to individual strategies as our fitness function, we instead use the size of their basin of attraction under an evolutionary game-theoretic analysis. We apply our method to the double auction market [6]: a domain which has long served as a benchmark problem in understanding strategic-interactions in multi-agent systems.

The outline of this paper is as follows. In Sections 2 and 3 we give a brief introduction to game theory and evolutionary game theory within the context of automated strategy-acquisition. In Section 4 we review some of the weaknesses of standard co-evolutionary algorithms from the perspective of game-theory. In Section 5 we give a brief overview of a technique called heuristic-strategy approximation, also known as empirical game theory, on which our method is based, and discuss how it is able to overcome some of the limitations of co-evolutionary algorithms when analysing strategic interactions amongst *existing* well-known strategies for a given game. In Section 6 we perform a heuristic-strategy analysis on a variant of the double auction market called a clearing-house. In Section 7 we describe our novel contribution to this domain; viz, how we are able to combine heuristic-strategy analysis and evolutionary search in order to automatically acquire *new* strategies for the double auction game, and finally in sections 8 and 9 we discuss our results.

2. NASH EQUILIBRIUM

The failure of certain types of co-evolutionary algorithms to converge on robust strategies in certain scenarios is well known, and has many possible causes; for example, the population may enter a limit cycle if strategies learnt in earlier generations are able to exploit current opponents and current opponents have "forgotten" how to beat the revived living fossil. Whilst many effective techniques have been developed to overcome these problems, there remains, however, a deeper problem which is only beginning to be addressed successfully. In some games, such as Chess, we can safely bet that if player A consistently beats player B, and player B consistently beats player C, then player A is likely to beat

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.

player C. Since the dominance relationship is transitive, we can build meaningful *rating systems* for objectively ranking players in terms of ability, and the use of such ranking systems can be used to assess the "external" fitness of strategies evolved using a co-evolutionary process and ensure that the population is evolving toward better and better strategies. In many other games, however, the dominance graph is highly intransitive, making it impossible to rank strategies on a single scale. In such games, it makes little sense to talk about "best", or even "good", strategies since even if a given strategy beats a large number of opponent strategies there will always be many opponents that are able to beat it. The best strategy to play in such a game is always dependent on the strategies adopted by one's opponents.

Game theory provides us with a powerful concept for reasoning about the best strategy to adopt in such circumstances: the notion of a *Nash equilibrium*. A set of strategies for a given game is a Nash equilibrium if, and only if, no player can improve their payoff by unilaterally switching to an alternative strategy.

If there is no dominant strategy[1] for the game, then we should play the strategy that gives us the best payoff based on what we believe our opponents will play. If we assume our opponents are payoff maximisers, then we know that they will play a Nash strategy set by *reductio ad absurdum*; if they did not play Nash then by definition at least one of them could do better by changing their strategy, and hence they would not be maximising their payoff. This is very powerful concept, since although not every game has a dominant strategy, every finite game posseses at least one *equilibrium* solution [8]. Additionally, if we know the entire set of strategies and payoffs, we can deterministically compute the Nash strategies. If only a single equilibrium exists for a given game, this means that, in theory at least, we can always compute the "appropriate" strategy for a given game.

Note, however, that the Nash strategy is not always the *best* strategy to play in all circumstances. For 2-player zero-sum games, one can show that the Nash strategy is not exploitable. However, if our opponents do not play their Nash strategy, then there may be other non-Nash strategies that are better at exploiting off-equilibrium players. Additionally, many equilibria may exist and in n-player non-constant-sum games it may be necessary for agents to *coordinate* on the same equilibrium if their strategy is to remain secure against exploitation; if we were to play a Nash strategy from one equilibrium whilst our opponents play a strategy from an alternative equilibrium we may well find that our payoff is significantly lower than if we had coordinated on the same equilibrium as our opponents.

3. BEYOND NASH EQUILIBRIUM

Standard game theory does not tell us which of the many possible Nash strategies our opponents are likely to play. *Evolutionary* game theory [11] and its variants attack this problem by positing that, rather than computing the Nash strategies for a game using brute-force and then selecting one of these to play, our opponents are more likely to gradually adjust their strategy over time in response to to repeated observations of their own and others' payoffs. One approach to evolutionary game-theory uses the *replicator dynamics* equation to specify the frequency with which different pure strategies should be played depending on our opponent's strategy:

$$\dot{m}_j = [u(e_j, \vec{m}) - u(\vec{m}, \vec{m})]\, m_j \tag{1}$$

where $\vec{m}$ is a mixed-strategy vector, $u(\vec{m}, \vec{m})$ is the mean payoff when all players play $\vec{m}$, and $u(e_j, \vec{m})$ is the average payoff to pure strategy j when all players play $\vec{m}$, and $\dot{m}_j$ is the first derivative of m_j with respect to time. Strategies that gain above-average payoff become more likely to be played, and this equation models a simple *co-evolutionary* process of mimicry learning, in which agents switch to strategies that appear to be more successful.

For any initial mixed-strategy we can find the eventual outcome from this co-evolutionary process by solving $\dot{m}_j = 0$ for all j to find the final mixed-strategy of the converged population. This model has the attractive properties that: (i) all Nash equilibria of the game are stationary points under the replicator dynamics; and (ii) all focal points of the replicator dynamics are Nash equilibria of the evolutionary game.

Thus the Nash equilibrium solutions are embedded in the stationary points of the direction field of the dynamics specified by equation 1. Although not all stationary points are Nash equilibria, by overlaying a dynamic model of learning on the equilibria we can see which solutions are more likely to be discovered by *boundedly-rational* agents. Those Nash equilibria that are stationary points at which a larger range of initial states will end up, are equilibria that are more likely to be reached (assuming an initial distribution that is uniform).

This is all well and good in theory, but the model is of limited practical use since many interesting real-world games are *multi-state*[2]. Such games can be transformed into normal-form games, but only by introducing an intractably large number of pure strategies, making the payoff matrix impossible to compute.

4. CO-EVOLUTION

But what if we were to approximate the replicator dynamics by using an evolutionary search over the strategy space? Rather than considering an infinite population consisting of a mixture of all possible pure strategies, we use a small finite population of randomly sampled strategies to approximate the game. By introducing mutation and cross-over, we can search hitherto unexplored regions of the strategy space. Might such a process converge to some kind of approximation of a true Nash equilibrium? Indeed, this is one way of interpreting existing co-evolutionary algorithms; fitness-proportionate selection plays a similar role to the replicator dynamics equation in ensuring that successful strategies propagate, and genetic operators allow them to search over novel sets of strategies. There are a number of problems with such approaches from a game-theoretic perspective, however, which we shall discuss in turn.

Firstly, the proportion of the population playing different strategies serves a dual role in a co-evolutionary algorithm [4]. On the one hand, the proportion of the population playing a given strategy represents the probability of playing

[1] A strategy which is always the best one to adopt no matter what any opponent does.

[2] The payoff for a given move at any stage of the game depends on the history of play.

that pure strategy in a mixed-strategy Nash equilibrium. On the other hand, evolutionary search requires diversity in the population in order to be effective. This suggests that if we are searching for Nash equilibria involving mixed-strategies where one of the pure strategy components has a high frequency, corresponding to a co-evolutionary search where a high percentage of the population is adopting the same strategy, then we may be in danger of over-constraining our search as we approach a solution.

Secondly and relatedly, although the final set of strategies in the converged population may be best responses to each other, there is no guarantee that the final mix of strategies cannot be invaded by other yet-to-be-countered strategies in the search space, or strategies that became extinct in earlier generations because they performed poorly against an earlier strategy mix that differed from the final converged strategy mix. Genetic operators such as mutation or cross-over will be poor at searching for novel strategies that could potentially invade the newly established equilibrium because of the above problem. If these conditions hold, then the final mix of strategies is implausible as a true Nash equilibrium or ESS, since there will be unsearched strategies that could potentially break the equilibrium by obtaining better payoffs for certain players. We might, nevertheless, be satisfied with the final mix of strategies as an approximation to a true Nash equilibrium on the grounds that if our co-evolutionary algorithm is unable to find equilibrium-breaking strategies, then no other algorithm will be able to do so. However, as discussed above, we expect *a priori* that co-evolutionary algorithms will be particularly *poor* at searching for novel strategies once they have discovered a (partial) equilibrium.

Finally, co-evolutionary algorithms employ a number of different selection methods, not all of which yield population dynamics that converge on game-theoretic equilibria [5].

These problems have led researchers in co-evolutionary computing to design new algorithms employing game theoretic solution concepts [3]. In particular, [4] describe a game-theoretic search technique for acquiring approximations of Nash strategies in large symmetric 2-player constant-sum games with type-independent payoffs. In this paper, we address n-player non-constant-sum multi-state games with type-dependent payoffs. In such games, playing our Nash strategy (or an approximation thereof) does not guarantee us security against exploitation, thus if there are multiple equilibria, it may be more appropriate to play a *best-response* to the strategies that we infer are in play.

5. EMPIRICAL GAME-THEORY

Reeves et al. [1] and Walsh et al. [12] obviate many of the problems of standard co-evolutionary algorithms by restricting attention to small representative sample of "heuristic" strategies that are known to be commonly played in a given multi-state game. For many games, unsurprisingly none of the strategies commonly in use is dominant over the others. Given the lack of a dominant strategy, it is then natural to ask if there are mixtures of these "pure" strategies that constitute game-theoretic equilibria.

For small numbers of players and heuristic strategies, we can construct a relatively small normal-form payoff matrix which is amenable to game-theoretic analysis. This *heuristic* payoff matrix is calibrated by running many iterations of the game; variations in payoffs due to different player-types (eg black or white, buyer or seller) or stochastic environmental factors (eg PRNG seed) are averaged over many samples of type information resulting in a single mean payoff to each player for each entry in the payoff matrix. Players' types are assumed to be drawn independently from the same distribution, and an agent's choice of strategy is assumed to be independent of its type, which allows the payoff matrix to be further compressed, since we simply need to specify the number of agents playing each strategy to determine the expected payoff to each agent. Thus for a game with k strategies, we represent entries in the heuristic payoff matrix as vectors of the form

$$\vec{p} = (p_1, \ldots p_k)$$

where p_i specifies the number of agents who are playing the ith strategy. Each entry $p \in P$ is mapped onto an outcome vector $q \in Q$ of the form

$$\vec{q} = (q_1, \ldots q_k)$$

where q_i specifies the expected payoff to the ith strategy. For a game with n agents, the number of entries in the payoff matrix is given by

$$s = \frac{(n+k-1)!}{n!(k-1)!} \tag{2}$$

For small n and small k this results in payoff matrices of manageable size; for $k = 3$ and $n = 6$, 8, and 10 we have $s = 28$, 45, and 66 respectively.

Once the payoff matrix has been computed we can subject it to a rigorous game-theoretic analysis, search for Nash equilibria solutions and apply different models of learning and evolution, such as the replicator dynamics model, in order to analyse the dynamics of adjustment to equilibrium.

The equilibria solutions that are thus obtained are not rigorous Nash equilibria for the full multi-state game; there is always the possibility that an unconsidered strategy could invade the equilibrium. Nevertheless, heuristic strategy equilibria are more plausible as models of real world game playing than those obtained using a standard co-evolutionary search precisely because they *restrict* attention to strategies that are commonly known and are in common use. We can therefore be confident that no commonly known strategy for the game at hand will break our equilibrium, and thus the equilibrium stands at least some chance of persisting in the short term future.

Of course, once an equilibrium is established, the designers of a particular strategy may not be satisfied with their strategy's adoption-rate in the game-playing population at large. As [12] suggest, the designers of, for example, a particular trading strategy in a market game may have financial incentives such as patent rights to increase their "market-share" – that is, the proportion of players using their strategy, or, in game-theoretic terms, the probability of their pure strategy being played in a mixed-strategy equilibrium with a large basin of attraction. They go on to propose a simple methodology for performing such optimization using manual design methods. A promising-looking candidate strategy is chosen for perturbation analysis; a new, perturbed, version of the original heuristic payoff matrix is computed in which the payoffs of the candidate strategy are increased by a small fixed percentage, thus modelling a hypothetical tweak to the strategy that yields in a small increase in payoffs. The replicator-dynamics direction field is then replotted to establish whether the hypothetically-optimized strategy is able to

to achieve a high adoption rate in the population. Strategy designers can then concentrate their efforts on improving those strategies that become strong attractors with a small increase in payoffs.

In this paper, we extend this technique by using a genetic-algorithm (GA) to *automatically* optimize candidate strategies by searching for a hitherto-unknown best-response — or, to use more appropriate nomenclature, a *better-response* — to an existing mix of heuristic strategies. Rather than using a standard co-evolutionary algorithm to perform the optimization, we use a single-population GA where the fitness of an individual strategy is computed from the heuristic-strategy payoff matrix according to the basin size it yields under the replicator dynamics.

6. THE DOUBLE-AUCTION

We apply our method to the acquisition of strategies for the *double auction* [6]. The double auction is a generalisation of more commonly-known single-sided auctions, such as the English ascending auction, which involve a single seller trading with multiple buyers. In a double auction, we allow multiple traders on both sides of the market; as well as soliciting offers to buy a good from buyers, that is bids, we also solicit offers to sell a good from sellers, so called *asks*. Variants of the double auction are commonly used in many real-world market places such as stock exchanges in scenarios where supply and demand are highly dynamic. Whilst single-sided auctions are well-understood from a game theoretic perspective, double-sided auctions remain intractable to a full game-theoretic analysis especially when there are relatively few traders on each side of the market. Thus much analysis of this game has focused on using heuristic methods to explore viable bidding strategies.

In previous work [10], we used a heuristic-strategy analysis to analyse two variants of the double auction market mechanism populated with a mix of heuristic strategies, and were able to find approximate game-theoretic equilibrium solutions. In this paper, we use the same basic framework, but we focus on the *clearing-house* double auction (CH) [6] with uniform pricing, in which all agents are polled for their offers before transactions take place, and all transactions are then executed at the same market-clearing price. In this paper, we consider only the following three heuristic-strategies. Future work will use a more representative (and larger) set of heuristic-strategies to optimize against.

- The truth-telling strategy (TT), whereby agents submit offers equal to their valuation for the resource being traded (in a strategy-proof market, TT will be a dominant strategy);
- The Roth-Erev strategy (RE) – a strategy based on reinforcement learning, described in [2] and calibrated with the parameters specified in [9]; and
- The Gjerstad-Dickhaut strategy (GD) [7], whereby agents estimate the probability of any bid being accepted based on historical market data and then bid to maximize expected profit.

Since all mixed-strategy vectors lie in the unit-simplex, for $k = 3$ strategies we can project the unit-simplex onto a two dimensional space and then plot the switching between strategies that occurs under the dynamics of equation 2. Figure 1 shows the direction-field of the replicator-dynamics equation for these three heuristic strategies, showing that we have two equilibrium solutions. Firstly, we see that GD is a best-response to itself, and hence is a pure-strategy equilibrium. We also see it has a very large *basin of attraction*; for any randomly-sampled initial configuration of the population most of the flows end up in the bottom-right-hand-corner. Additionally, there is a second mixed-strategy equilibria at the coordinates (0.88, 0.12, 0) in the simplex corresponding to an 88% mix of TT and a 12% mix of RE, however the attractor for this equilibrium is much smaller than the pure-strategy GD equilibrium; only 6% of random starts terminate here vs 94% for pure GD. Hence, according to this analysis, we expect most of the population of traders to adopt the GD strategy.

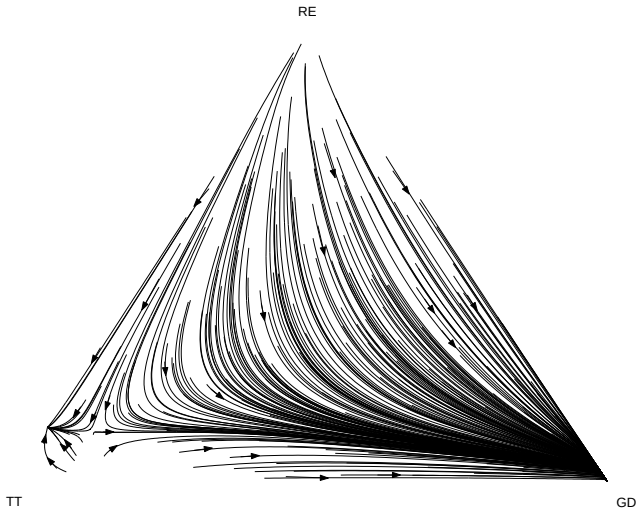

Figure 1: The original replicator dynamics direction field for a 12-agent clearing-house auction with the original unoptimized Roth-Erev strategy (labelled RE).

How much confidence can we give to this analysis given that the payoffs used to construct the direction-field plot were estimated based on only 2×10^3 samples of each game? One approach to answering this question is to conduct a sensitivity analysis; we perturb the mean payoffs for each strategy in the matrix by a small percentage to see if our equilibria analysis is robust to errors in the payoff estimates. Figure 2 shows the direction-field plot after we perform a perturbation where we remove 2.5% of the payoffs from the TT and GD strategies and assign +5% payoffs to the RE strategy. This results in a qualitatively different set of equilibria; the RE strategy becomes a best-response to itself with a large basin of attraction (61%), and thus we conclude that our equilibrium analysis is sensitive to small errors in payoff estimates, and that our original prediction of widespread adoption of GD may not occur if we have underestimated the payoffs to RE.

If we observe a mixture of all three strategies in actual play, however, the perturbation analysis also suggests that we could bring about widespread defection to RE if were able to tweak the strategy by improving its payoff slightly; *the perturbation analysis points to RE as a candidate for potential optimization.*

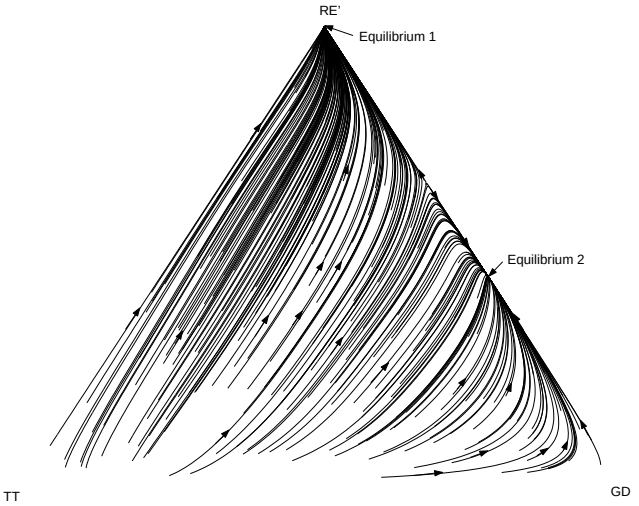

Figure 2: Replicator dynamics direction field for a 12-agent clearing-house auction perturbed with +5% payoffs to the Roth-Erev strategy (labeled RE')

7. STRATEGY ACQUISITION

In the previous section we described how we used heuristic-strategy approximation to identify a candidate strategy for optimization. We also introduced an intriguing metric for ranking strategies on a single fully-ordered scale: viz, the size of the strategy's basin of attraction under the replicator dynamics. In this section we use this metric to perform a heuristic search of a space of strategies closely related to the RE strategy. In the following we define the space of strategies that we search, and the details of the search algorithm.

The RE strategy uses re-inforcement learning (RL) to choose from n possible markups over the agent's limit price based on a reward signal computed as a function of profits earned in the previous round of bidding. Agents bid or ask at price p

$$p = l \pm mo \tag{3}$$

where l is the agent's limit price, o is the output from the learning algorithm and m is a scaling parameter. Additionally, the Roth-Erev learning algorithm itself has several free parameters: the recency parameter r, the experimentation parameter x, and an initialisation parameter $s1$. In addition to the original Roth-Erev (RE) algorithm, there are several other learning-algorithms that that have successfully been used for RL strategies in ACE. We search over three additional possibilities: stateless Q-learning (SQ), modifications to RE used by [9] (NPT) and a control algorithm which selects a uniformly random action regardless of reward signal (DR). SQ has free parameters: the discount-rate g, experimentation ϵ, and a learning-rate p.

Individuals in this search space were represented as a 50-bit string, where:

- bits 1-8 coded for parameter m in the range $(1, 10)$;
- bits 9-16 coded for the parameters ϵ or x in the range $(0, 1)$;
- bits 17-24 coded for parameter n in the range $(2, 258)$;
- bits 25-32 coded for parameters g or r in the range $(0, 1)$;
- bits 33-40 coded for parameter $s1$ in the range $(1, 1.5 \times 10^4)$;
- bits 41-42 coded for the choice of learning algorithm amongst RE, NPT, SQ or DR; and
- Bits 43-50 coded for parameter p in the range $(0, 1)$.

We used a genetic-algorithm (GA) to search this space of strategies, where the fitness of each individual strategy in the search space was computed by estimating its basin size under the replicator dynamics under interaction with our existing three strategies: GD, TT and RE. Basin size was estimated using the same brute-force methods described in Section 5, but since we recompute all entries in the heuristic-payoff matrix in support of each candidate strategy, we used lower sample sizes in order to facilitate evaluation of many strategies; the sample size for the number of games played for each entry in the heuristic payoff matrix was increased as a function of the generation number: $10 + int(100 \ln(g + 1))$ allowing the search-algorithm to quickly find high-fitness regions of the search-space in earlier generations and reducing noise and allowing more refinement of solutions in later generations. We used a constant number of replicator-dynamics trajectories — 50 — in order to estimate the basin size from the payoff matrix once it had been recomputed for our candidate strategy.

We chose a GA as our search algorithm principally because of its ability to cope with the additional noise that the lower sample size introduced into the objective function Our GA was configured with a population size of 100, with single-point cross-over, a cross-over rate of 1, a mutation-rate of 10^{-4} and fitness-proportionate selection. The GA was run for 32 generations, which took approximately 1800 CPU hours.

8. RESULTS

Figure 5 shows the mean fitness of the GA population for each generation. As can be seen, there is still a large amount of variance in fitness values in later generations. However, inspection of a random sample of strategies from each generation revealed a partial convergence of phenotype, but with significant fluctuations in fitness values due to small sample sizes (see above). Most notably, the fittest individual at generation 32 had also appeared intermittently as the fittest individual five times in the previous 10 generations, and thus we took this as the output from the search.

Our optimised strategy uses the stateless Q-learning algorithm with the following parameters:

$$\begin{aligned} m &= 1.210937 \\ n &= 6 \\ \epsilon &= 0.18359375 \\ g &= 0.4140625 \\ p &= 0.1875 \end{aligned}$$

The notable feature of this strategy is the small number of possible markups (n), and the narrow range of the markups

$[1, nm]$ as compared with the distribution of valuation distribution widths. This feature was shared by all of the top five strategies in the last ten generations, and is another factor that indicated convergence of the search.

We proceeded to analyze our specimen strategy under a full heuristic-strategy analysis using 10^4 samples of the game for each of the 455 entries in the payoff matrix. With the current version of our simulator[3], we are able to complete this analysis in less than twenty four hours using a dual-processor 3.6Ghz Xeon workstation.

Figure 6 shows twenty trajectories of the replicator dynamics plotted as a time series for each strategy, and shows the interaction between the new, optimised strategy, OS, together with the existing strategies: GD, TT and RE.

Taking 10^3 randomly sampled initial mixed-strategies, we calculate that there are two attractors:

$$\begin{aligned} A &= (0.0, 0.0, 1.0, 0.0) \\ B &= (0.67, 0.32, 0.0, 0.0) \end{aligned}$$

over (OS, TT, GD, RE). Attractor A captures only 3% of trajectories, whereas attractor B captures virtually the entire four-dimensional simplex (97%). Although this basin is very large, our optimized strategy shares this equilibrium with the truth-telling strategy (TT), giving us a final total market share to $0.67 \times 0.97 = 65\%$. This compares favourably with a market-share of 32% for truth-telling and 3% for GD. The original RE strategy is dominated by our optimised strategy. Figures 3 and 4 show the direction field for two of the 3-strategy combinations involving our optimised strategy: (OS, TT, GD) and (OS, GD, RE) respectively.

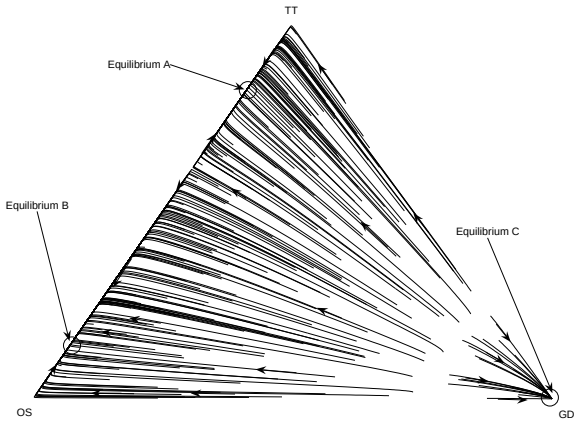

Figure 3: Replicator dynamics direction field for a 12-agent clearing-house auction showing interaction between optimised strategy (OS) verses TT and GD

9. DISCUSSION

It is somewhat remarkable that our somewhat simplistic optimised strategy is able to gain defectors from a highly sophisticated strategy like GD, whilst at the same time truth-telling is able to retain a share of followers in a population predominated by *OS*ers (TT appears to be *parasitic* on OS). What accounts for the ability of small OS mixes to invade high-probability mixes of a sophisticated adaptive strategy (GD), whilst remaining vulnerable to invasion by a low-probability mix of a non-adaptive strategy (TT)?

Our current hypothesis is that OS is able to exploit a flaw in the way that we construct valuation distributions. As discussed earlier, we use the same method of assigning valuations as in [12]; that is, for each run of the game, the lower-bound, b, of the valuation distribution is selected uniformly at random from the range $[61, 160]$ and the upper-bound b' is similarly drawn from $[b + 60, b + 209]$. For that run of the game, each agent's valuation is then drawn uniformly from $[b, b']$. However, we hypothesise that this results in a statistical correlation between the meta-bounds and the average slope of truthful supply and demand schedules— that is, given these distribution parameters there is insufficient variance in the difference between valuations of traders who are neighbours on the supply or demand curve. Since we are using a uniform-price k=0.5 clearing rule, the mechanism is vulnerable to price-manipulation from the least efficient trades; the buyer with the lowest matched bid, and the seller with the highest matched ask can potential manipulate the final clearing price - *provided that they do not overstate their value claim to the extent that it impinges on the 2nd-lowest matched bid, or the 2nd-highest matched ask*. Given sufficient variance in valuations between the lowest and the 2nd-lowest matched offers (call this random variable the "match delta"), this vulnerability is not easily exploited. However, in a market with a small number of traders and a narrow distribution for the match delta there is an opportunity to trade at small margin above truth if you find yourself with a valuation close to the equilibrium price. This is precisely the behaviour of the strategies that we observe to be predominant in the later generations of our GA- they all use a small number of possible markups, each of them small in comparison to the possible valuation bounds. The reinforcement-learning component of the strategy is then able to fine-tune the markup depending on where the trader finds themselves on the supply or demand curve after valuations are drawn. If it is far away from the equilibrium-price it can adjust its margin close to zero, whilst if it is near the equilibrium-price it can find a small margin that does not impinge on its nearest-neighbour. This hypothesis is also consistent with parasitic truth-telling; it is easy to see that truth-telling is a best-response for a 2nd-lowest matched bidder to a lowest matched bidder playing OS. Further work needs to be carried out to verify this hypothesis, but early indications are that the efficacy of our optimised strategy is highly sensitive to changes in the valuation meta-bounds.

Thus it appears that our evolved strategy over-fits to what is potentially an artifact. Does this invalidate our approach? On the contrary, we believe that it demonstrates that highly sophisticated strategies can possess an Achilles heel in certain situations, and that our method for strategy acquisition can find and exploit these vulnerabilities. GD performs robustly given a large variance in match delta, but given some partial prior knowledge about valuation distributions in a small market, then there may exist a more effective, and simpler strategy. In the particular case that valuation distributions are constructed as above, then the strategy OS will perform remarkably well.

10. SUMMARY

[3] `http://freshmeat.net/projects/jasa`

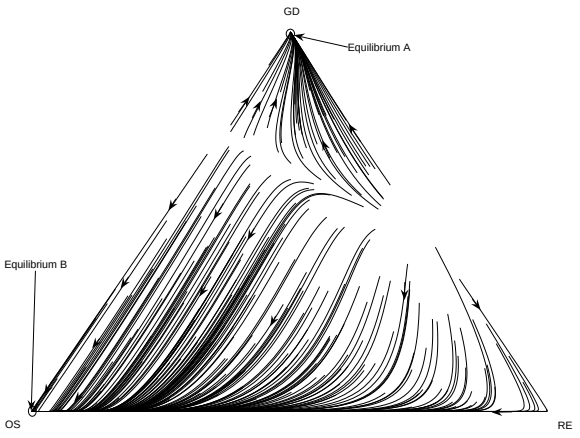

Figure 4: Replicator dynamics direction field for a 12-agent clearing-house auction showing interaction between optimised strategy (OS) verses GD and the original Roth-Erev strategy (RE)

In this paper we have described a novel method for strategy acquisition for n-player non-zero-sum games, and we have successfully applied our method to the acquisition of a trading strategy for the double-auction market; a recognised benchmark problem for this domain. Our method combines two existing methodologies in a very simple way: empirical game-theory and heuristic search (in this case, using a genetic-algorithm). Our key contribution to this domain is the use of the basin-size metric to rank strategies. Although we have not provided a thorough analytical verification of our method, we believe that at this stage its novel features over and above those of empirical game-theory are simple enough that a successful application of our method to well-known benchmark problem suffices to demonstrate the potential for future research.

11. FURTHER WORK

The main weaknesses of our current approach stem from: a) potential weaknesses in the underlying empirical game-theory methodology, and b) the lack of a customised search algorithm tailored specifically for this domain.

As regards a), we recognise that empirical game-theory is an emerging field. As such, advances and refinements are continually being made to address some of its earlier weaknesses. For example, Walsh et al. [13] describe a refinement to the method which allows for a more principled approach for determining optimal sample size, and [14] describe a further approximation technique that allows for a greatly reduced number of strategy profiles to be considered. These and other refinements will be incorporated into our approach in future work. One of the key acknowledged weaknesses of empirical game-theory is the somewhat arbitrary selection of initial heuristic-strategies. However, it is our hope that the search method outlined in this paper will become the basis of an iterative approach to extending an initial set of manually chosen heuristic-strategies by populating the mix with additional heuristic-strategies that are discovered through heuristic-search.

As regards b), designing such a search algorithm is future work. But note that *despite* the lack of a custom-designed search algorithm, we are still able to use a general-purpose heuristic-search method to successfully acquire an interesting novel strategy for our benchmark problem.

12. ADDITIONAL AUTHORS

Additional author: P. McBurney (Department of Computer Science, University of Liverpool,
email: `peter@csc.liv.ac.uk`).

13. REFERENCES

[1] M. W. D.M. Reeves, J.K. MacKie-Mason and A. Osepayshvili. Exploring bidding strategies for market-based shceduling. *Decision Support Systems*, 2005.

[2] I. Erev and A. E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 88(4):848–881, 1998.

[3] S. Ficici. *Solution Concepts in Coevolutionary Algorithms*. PhD thesis, Brandeis University, May 2004.

[4] S. Ficici and J. Pollack. A game-theoretic memory mechanism for coevolution. In *LNCS 2723*, pages 286–297. Springer-Verlag, 2003.

[5] S. G. Ficici and J. B. Pollack. A game-theoretic approach to the simple coevolutionary algorithm. In *PPSN VI*. Springer Verlag, 16-20 2000.

[6] D. Friedman and J. Rust. *The Double Auction Market*. Westview, 1991.

[7] S. Gjerstad and J. Dickhaut. Price formation in double auctions. *Games and Economic Behaviour*, 22:1–29, 1998.

[8] J. Nash. Equilibrium points in n-person games. In *Proc. of the National Academy of Sciences*, volume 36, pages 48–49, 1950.

[9] J. Nicolaisen, V. Petrov, and L. Tesfatsion. Market power and efficiency in a computational electricity market with discriminatory double-auction pricing. *IEEE Trans. on Evol. Computation*, 5(5):504–523, October 2001.

[10] S. Phelps, S. Parsons, and P. McBurney. An evolutionary game-theoretic comparision of two double auction market designs. In *Agent-Mediated Electronic Commerce VI*. Springer-Verlag, 2004.

[11] J. Smith. *Evolution and the theory of games*. Cambridge University Press, 1982.

[12] W. Walsh, R. Das, G. Tesauro, and J. Kephart. Analyzing complex strategic interactions in multi-agent games. In *AAAI-02 Workshop on Game Theoretic and Decision Theoretic Agents*, 2002.

[13] W. Walsh, D. Parkes, and R. Das. Choosing samples to compute heuristic-strategy nash equilibrium. In *Agent-Mediated Electronic Commerce V*, 2003.

[14] M. P. Wellman, D. M. Reeves, K. M. Lockner, and R. Suri. Searching for walverine. In *Proceedings of the IJCAI-05 Workshop on Trading Agent Design and Analysis*, 2005.

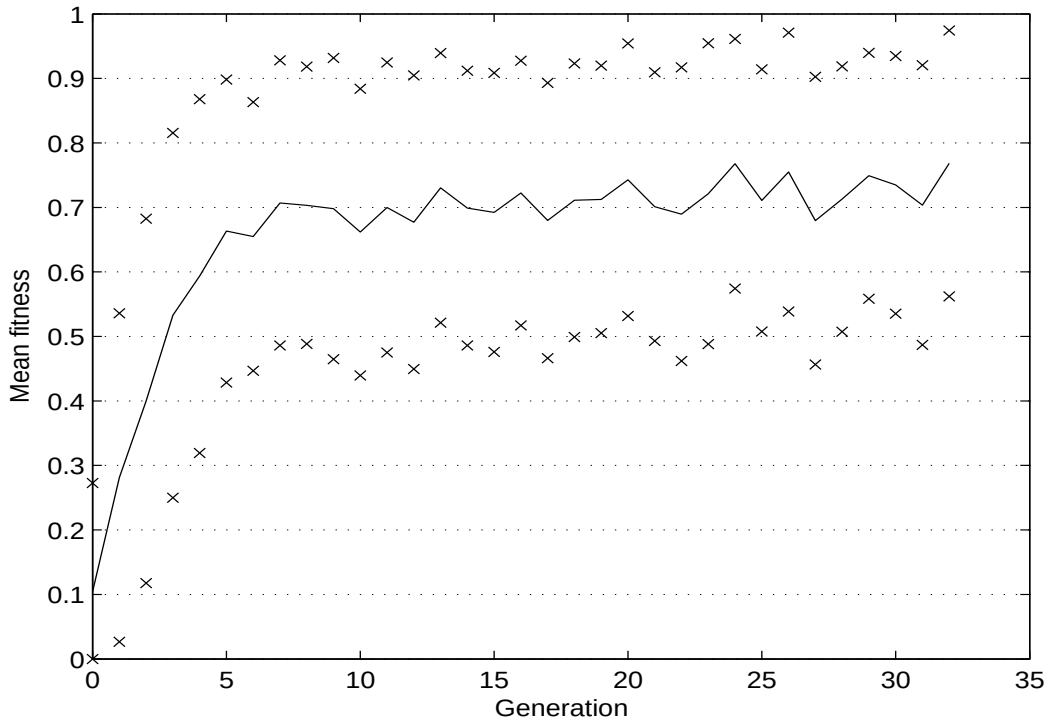

Figure 5: Mean fitness of the GA population with one stdev

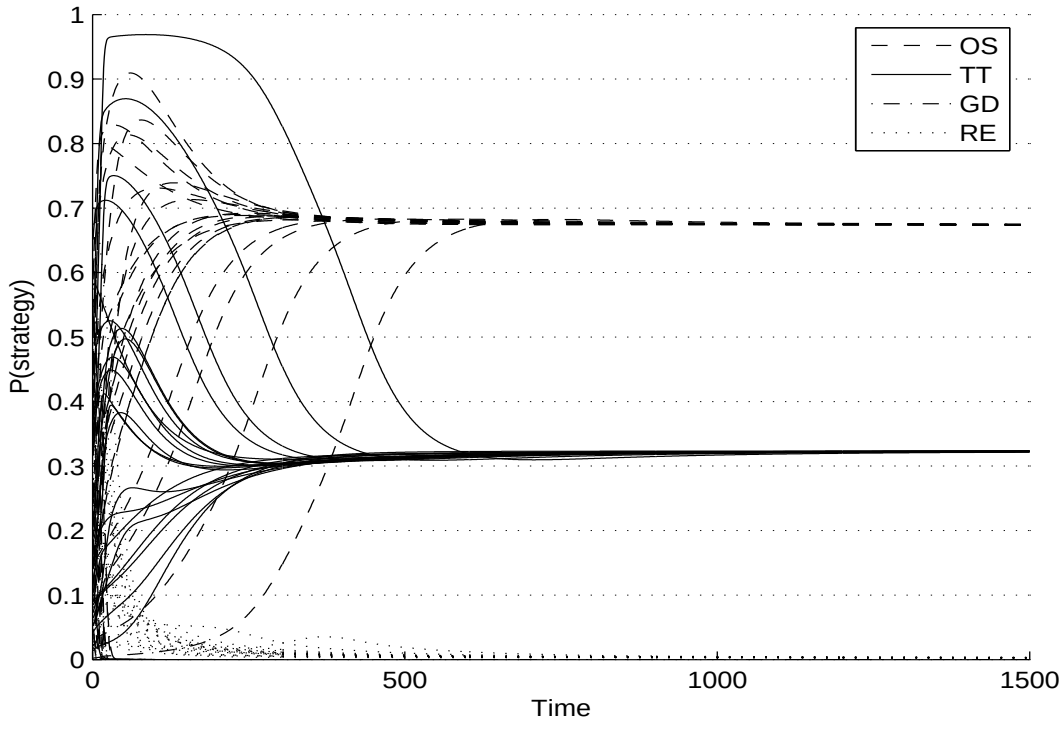

Figure 6: Replicator dynamics time series plot for a 12-agent clearing-house auction showing interaction between optimised strategy (OS) verses GD, TT and the original Roth-Erev strategy (RE)