Using Population-based Search and Evolutionary Game Theory to Acquire Better-response Strategies for the Double-Auction Market

S. Phelps
sphelps@csc.liv.ac.uk

M. Marcinkiewicz
mom7@cs.columbia.edu

S. Parsons
parsons@sci.brooklyn.cuny.edu

P. McBurney
peter@csc.liv.c.uk

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.

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. The reasons for this are many and varied; 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 to 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 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 [Nash, 1950]. 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 equi-

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

libria 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.

2 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 [Smith, 1982] 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:

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

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 attractors 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.

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 [Ficici and Pollack, 2003]. On the one hand, the proportion of the population playing a given strategy represents the probability of playing 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.

Thirdly, in the case where there are multiple equilibria, the particular one to which our population converges will be highly sensitive to the initial configuration of the population, that is the particular mix of random strategies that we start with, and certain equilibrium solutions may only be obtainable if we start with a given mix of initial strategies. In evolutionary game theory, we can simply take many samples of initial mixed-strategy vectors and for each of them solve the replicator dynamics equation in order to find stationary points. However, such brute-force approaches require

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

the sampling of many thousands of initial mixed strategies in order to accurately assess the population dynamics of a three-strategy game. If we translate this into a co-evolutionary algorithm with a large strategy space, it necessitates running the co-evolutionary process hundreds of thousands of times with different randomly initialised populations in order to discover *robust* equilibria, which is computationally impractical in most cases.

Finally, co-evolutionary algorithms employ a number of different selection methods, not all of which yield population dynamics that converge on game-theoretic equilibria [Ficici and Pollack, 2000].

These problems have led researchers in co-evolutionary computing to design new algorithms employing game-theoretic solution concepts [Ficici, 2004]. In particular, [Ficici and Pollack, 2003] 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.

3 Heuristic-strategy approximation

[Walsh *et al.*, 2002] obviate many of the problems of standard co-evolutionary algorithms by restricting attention to a 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:

$$\frac{(n+k-1)!}{n!(k-1)!}$$

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 co-evolutionary search*[3] precisely because they restrict attention to strategies that are commonly known and are in common use. Assuming that we have incorporated all commonly known strategies into our analysis, we can 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 [Walsh *et al.*, 2002] 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

[3]Note that we are not claiming that heuristic-strategy equilibria are more plausible than equilibria obtained from a full (but unfeasible) game-theoretic analysis.

optimization, we use a single-population GA where the fitness of an individual strategy is computed from the heuristic-strategy payoff matrix according to its expected payoff when it is played against the existing mixed strategy.

4 An HSA analysis of a double-auction

We apply our method to the acquisition of strategies for the *double-auction* [Friedman and Rust, 1991]. 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 *agent-based computational economics* (ACE) [Tesfatsion, 2002] to explore viable bidding strategies.

[Phelps *et al.*, 2004] 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) [Friedman and Rust, 1991] 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 preliminary work, we consider only the following three heuristic-strategies:[4]

- 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 myopic reinforcement-learning in which agents increase their propensity to bid at a particular markup based on their profits earnt in the previous round, described in [Nicolaisen *et al.*, 2001] and calibrated as specified therein; and
- The Gjerstad-Dickhaut strategy (GD) [Gjerstad and Dickhaut, 1998], 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 1. 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.

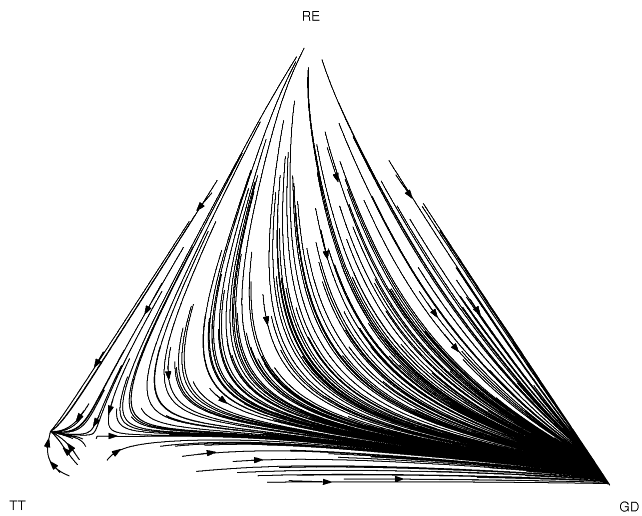

Figure 1: The original replicator dynamics direction field for a 12-agent clearing-house auction with the original unoptimized reinforcement-learning strategy (labeled RE).

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 2000 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 optimiza-*

[4]Future work will use a more representative (and larger) set of heuristic-strategies to optimize against.

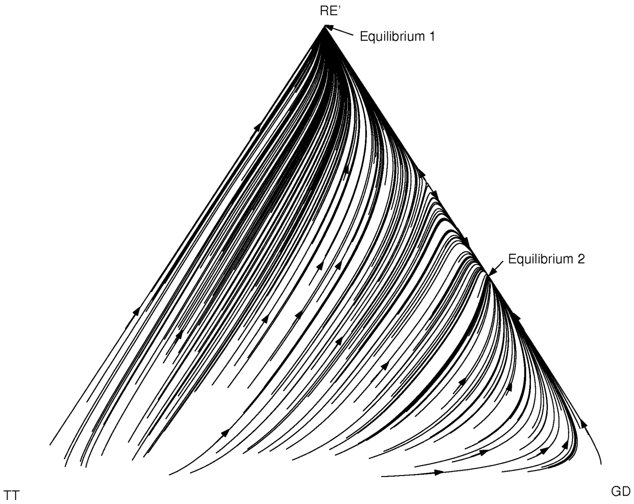

Figure 2: Replicator dynamics direction field for a 12-agent clearing-house auction perturbed with +5% payoffs to the reinforcement-learning strategy (labeled RE')

tion. In this paper, we describe a method for performing this optimization automatically by using a genetic algorithm.

5 Optimizing RE

Our design goal is to bring about widespread defection to our optimised strategy by increasing the size of its basin of attraction. This immediately suggests a possible fitness function; we could estimate the size of the resulting basin for each strategy by solving equation 1 for a given sample of initial random mixed strategies. However, performing this computation with sufficiently large sample size for accurate estimates is too expensive to be practical.

As an alternative, we use a heuristic metric. Our optimised strategy will be able to win defectors from a given starting point if it is able to obtain higher than average payoffs when the mixed strategy represented by that point is being played, and for each subsequent point (mixed strategy) along the resulting trajectory. Under the following simplifying assumptions: (i) each starting point is equally likely, (ii) trajectories do not overlap significantly, and (iii) the candidate strategy is indeed an attractor, our goal of maximising our basin of attraction corresponds to optimising payoffs to our candidate strategy for every possible mixture of heuristic strategies.

Since we know that the solution to the payoff-maximisation problem is likely to vary as we move around the unit simplex (for example, the best strategy to play in response to 100% truth-telling may vary from the best strategy to play against 100% GD), we should treat this as a multi-objective optimisation problem where each objective corresponds to the payoff to the candidate strategy in response to a given mixed strategy. More formally, given the (very large) set S of all possible pure strategies, a set $H = \{h_1, \ldots, h_k\} \subset S$ of heuristic strategies, the space of possible mixed heuristic strategies $\vec{m} = (m_1, \ldots, m_k) \in \triangle$, we can think of each $\vec{m} \in \triangle$ as an objective to optimize where the fitness of a solution, j, with respect to a given objective, $\vec{m}$, is given by the function $F : \triangle \times S \to \mathbf{R}$ where $F(\vec{m}, j) = u(e_j, \vec{m})$

However, given that: we have an infinite number of objectives; there is unlikely to be a dominant strategy; and the dominance graph is highly intransitive, we would expect the pareto-frontier of this multi-objective problem to contain an impractically large number of solutions[5]. We can combat this complexity by turning it into a single-objective problem; we weight each objective $\vec{m}$ equally, and define our fitness function as the average fitness across all objectives, viz. all mixed strategies in $\triangle$. This is formally equivalent to the expected payoff given that each mixed strategy is equally likely. We can simply express this as another mixed-strategy, since the payoff under a mixed strategy depends only on the probabilities expressed therein and the pure-strategy payoffs. Assuming that every $\vec{m} \in \triangle$ has equal probability, this in turn corresponds to the mixed strategy representing uniform probability over all pure strategies. Thus we want to find a strategy $s = \arg\max_{j \in S} F(\vec{c}, j)$ where $\vec{c} = (u_1, \ldots, u_k)$ and $u_i = \frac{1}{k}$

Note that since we will be competing with the original unoptimized version of the strategy we need to introduce an aditional heuristic strategy into the mix, yielding $k = 4$ strategies in total. Thus our fitness function is given by $F(\vec{c}, j)$ where $\vec{c} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ over the strategies TT, RE, GD and j.

This assumes that each starting point in the simplex is equally likely. If on the other hand, we had reason to believe that a particular mixed strategy was more likely to be in play (for example, by inferring probabilities based on the history of play in a manner akin to ficticious play), we might reweight our objectives correspondingly and search for a best response to a non-uniform mixed strategy.

5.1 Searching for a better-response

In order to optimize the RE strategy, we make its free parameters explicit and use this as our search space. One of parameters we consider is the choice of learning algorithm itself. Thus we perform a search over several myopic reinforcement learning algorithms and their associated free parameters.

The RE strategy uses reinforcement 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

$$p = l \pm mo \qquad (2)$$

where l is the agent's limit price, o is the output from the learning algorithm and m is a scaling parameter. The original version of the RE strategy uses the Roth-Erev learning algorithm [Erev and Roth, 1998] which has several free parameters: the recency parameter r, the experimentation parameter x, and an initialisation parameter $s1$.

In addition to the original Roth-Erev learning algorithm (ORE), 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

[5][Noble and Watson, 2001] report on results of a similar MOO approach for finding strategies for texas-holdem poker, in which the pareto frontier becomes unmanageably large.

(SQ), modifications to ORE used by [Nicolaisen *et al.*, 2001] (NPT) and a control algorithm which selects a uniformly random action regardless of reward signal (DR). SQ has free parameters: the discount-rate g, epsilon e, and a learning-rate p.

Individuals in the search space were represented as a 50-bit genome, where:

- bits 1-8 coded for parameter m in the range $(1, 10)$;
- bits 9-16 coded for the parameters e 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 $(1, 15000)$;
- bits 41-42 coded for the choice of learning algorithm amongst ORE, NPT, SQ or DR; and
- Bits 43-50 coded for parameter p in the range $(0, 1)$.

This space was searched using a GA 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 expected payoff to our candidate strategy was computed from the heuristic-strategy payoff matrix according to the fitness function $F(\vec{c}, j)$ where $\vec{c} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. Since one of our heuristic strategies is j, this necessitated recomputing all entries in the payoff matrix in support of j for each individual that was evaluated. We used a small number of samples of the game in order to populate each entry in the payoff matrix in the expectation that the GA would be robust to the additional noise that this would introduce into the payoffs; the sample size was increased as a function of the generation number: $10 + int(15 \ln(g + 1))$, thus allowing the GA 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.

6 Results

Figure 3 shows the mean fitness of the evolving population per generation. By generation 50, the population's mean fitness had plateaued to 0.94 with a standard error of 10^-3, and the estimated fitness of the best individual was 0.95 (based on 68 samples of the game). The best individual coded for a strategy using the stateless Q-learning algorithm with parameters $n = 54$, $m = 3.7421875$, $e = 0.0078125$, $g = 0.71875$ and $p = 0.21484375$.

The goal of this exercise was to see if we could find a replacement strategy for RE that would likely be adopted under replicator-dynamics learning given a population starting near the centre of the simplex. Figure 4 shows the direction-field of the replicator dynamics when we replace RE with our optimized strategy OS using 4900 samples of the game for each entry in the payoff matrix. As can be seen, our optimized strategy is a pure strategy equilibrium, and captures 51% of trajectories, compared with 47% for GD, and 2% for TT. Our new strategy is not as effective at invading *pure* TT

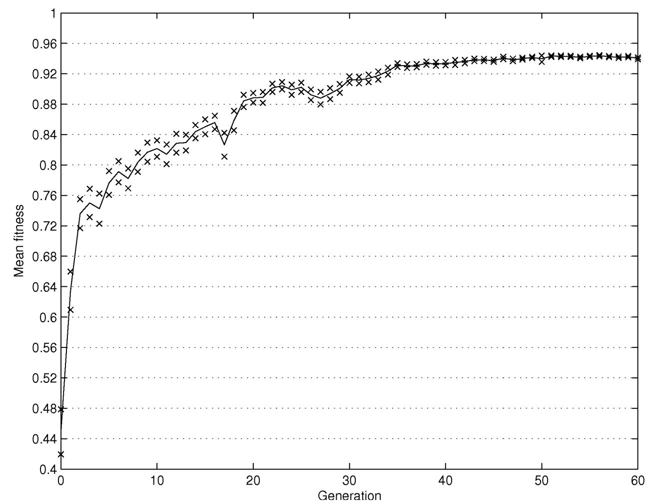

Figure 3: Mean fitness of the GA population with one standard error

as the original strategy, but it is effective at exploiting *high-frequency* TT populations, and thus is able to gain 100% defection to the optimized strategy instead of settling at a mixed TT/RE equilibrium[6].

However, the fact the original strategy performs better than the optimized strategy in this particular case implies that the original strategy is not dominated by our new strategy, and that it may be important to study the interaction between OS and RE. For example, it might be the case that OS is parasitic on RE, and is able to gain defectors by relying on RE to invade pure TT, and then in turn OS is able to invade an RE/TT mix. Work is underway to perform a full dynamic analysis of the 4-strategy game.

7 Conclusion

In this paper we have applied a novel method combining evolutionary search together with a principled game-theoretic analysis in order to automatically acquire a trading strategy for the double-auction market. We defined an appropriate measure of success in this game based on evolutionary game-theory, and we were able to demonstrate that our evolved strategy performed robustly according to this criterion.

Much existing work in this domain has used co-evolution to search for good all round strategies without regard to equilibrium or best-response considerations. However, we have argued that this is neither feasible nor desirable. Additionally, there is existing work which formulates equilibrium analysis as an optimization problem that can be solved using population-based search, for example, [Pavlidis *et al.*, 2005]. However, such approaches are applied to searching over mixed-strategies of closed normal-form games with

[6]Note that neither is GD able to exploit pure TT (hence TT becomes a best-response with a small basin of attraction), which is consistent with our original analysis. Note also, that in this case we are using a sealed-bid mechanism which may explain the strength of the truth-telling strategy.

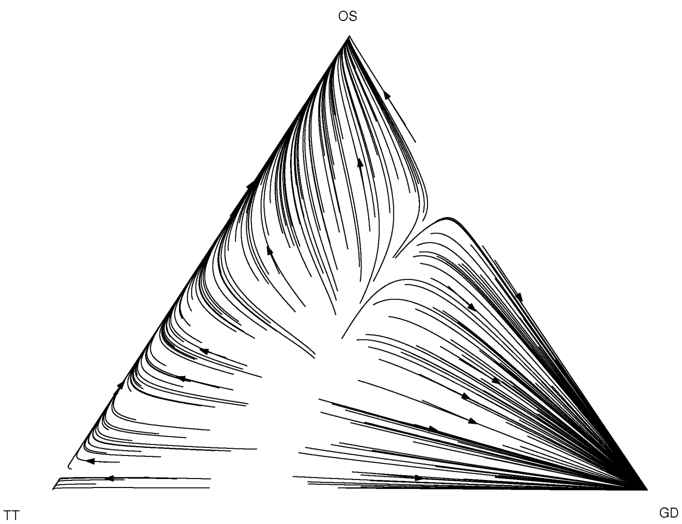

Figure 4: Replicator dynamics direction field for a 12-agent clearing-house auction using the GA-optimized strategy (labeled OS)

small numbers of pure strategies, and do not address the problem of searching for hitherto unconsidered pure strategies. To the best of our knowledge, our work is the first attempt to combine evolutionary search together with a principled game-theoretic analysis for acquisition of novel strategies in open-ended[7] n-player non-constant-sum multi-state games.

We recognize that much of this work is preliminary. In future work we will: extend the number of heuristic strategies that are analysed, incorporate the original strategy into the final dynamic analysis, use a more principled approach for calibrating the GA to deal with small sample sizes, search for better-responses to various non-uniform mixed heuristic strategies, use the technique iteratively to arrive at a more comprehensive set of heuristic-strategies (and thus a more accurate equilibrium analysis), and finally, we will use these extended heuristic-strategy equilibria in conjunction with the techniques in [Phelps *et al.*, 2004] in order to analyse the properties of different auctions from a mechanism design perspective.

References

[Erev and Roth, 1998] 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.

[Ficici and Pollack, 1998] S. G. Ficici and J. B. Pollack. Challenges in coevolutionary learning: Arms-race dynamics, open-endedness, and mediocre stable states. In *Proceedings of of ALIFE-6*, 1998.

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

[7]See [Ficici and Pollack, 1998] for a discussion of open-endedness

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

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

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

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

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

[Nicolaisen *et al.*, 2001] 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.

[Noble and Watson, 2001] J. Noble and R.A. Watson. Pareto coevolution: Using performance against coevolved opponents in a game as dimensions for pareto selection. In *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2001*, pages 493–50. Morgan Kauffman, 2001.

[Pavlidis *et al.*, 2005] N. G. Pavlidis, K.E. Parsopoulos, and M.N. Vrhatis'. Computing nash equilibria through computational intelligence methods. *Journal of Computational and Applied Mathematics*, 175:113–136, 2005.

[Phelps *et al.*, 2004] 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.

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

[Tesfatsion, 2002] L. Tesfatsion. Agent-based computational economics: growing economies from the bottom up. *Artificial Life*, 8(1):55–82, 2002.

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