Evolution of Market Mechanism Through a Continuous Space of Auction-Types III: Multiple Market Shocks Give Convergence Toward CDA

Dave Cliff
Information Infrastructure Laboratory
HP Laboratories Bristol
HPL-2002-312
November 27th , 2002*

E-mail: dave.cliff@hp.com

algorithmic mechanism design, auction and negotiation technology, automated trading, ZIP traders, genetic algorithms, e-marketplaces

This paper builds on previous papers describing our ongoing research in automated market-mechanism design: using a genetic algorithm (GA) to find optimal parameter-settings for software-agent traders that operate in virtual "e-marketplaces", where the rules of the marketplaces are *also* under simultaneous control of the GA. The aim is that the GA automatically designs new agent-based e-marketplaces that are more efficient than existing markets designed by (or populated by) humans. Das *et al.* recently demonstrated that ZIP software-agent traders consistently out-perform human traders in Continuous Double Auction (CDA) marketplaces similar to those used in the international financial markets. Cliff used a GA to explore a continuous space of ZIP-trader auction-market mechanisms, where the space of possible auction-types explored included the CDA and also two purely one-sided mechanisms. Surprisingly, the GA would sometimes settle on novel *hybrid* auction mechanisms partway between the CDA and a one-sided auction. Such results occurred when the market's supply and demand schedules were unchanging, and also when the schedules undergo a single sudden "shock" change halfway through the evaluation process. These results could *prima facie* support the hypothesis that hybrid auctions are in general preferable to the well-known CDA mechanism. In this paper we present new results that clarify that hypothesis. In our new experiments, more than one shock-change in supply/demand occurs during the mechanism-evaluation process; under this regimen, CDA auction-mechanisms are identified by the GA as optimal in three of the four experiments reported here, and in the fourth experiment a near-CDA hybrid mechanism was evolved. From these results we conclude that, while evolved hybrid market-mechanisms may be useful in niches where the marketplace's supply and demand dynamics are known a *priori* to be relatively stable or regular, the CDA remains the mechanism of choice when it is difficult or impossible to make accurate predictions concerning the dynamic stability of the supply and demand schedules.

* Internal Accession Date Only
Approved for External Publication

Evolution of Market Mechanism Through a Continuous Space of Auction-Types III: Multiple Market Shocks Give Convergence Toward CDA

Dave Cliff
Hewlett-Packard Laboratories Bristol
Filton Road, Bristol BS34 8QZ, U.K.
Tel. +44 (0) 117 312 8189
dave.cliff@hp.com

ABSTRACT

This paper builds on previous papers describing our ongoing research in automated market-mechanism design: using a genetic algorithm (GA) to find optimal parameter-settings for software-agent traders that operate in virtual "e-marketplaces", where the rules of the marketplaces are *also* under simultaneous control of the GA. The aim is that the GA automatically designs new agent-based e-marketplaces that are more efficient than existing markets designed by (or populated by) humans. Das *et al*. [8] recently demonstrated that ZIP software-agent traders consistently out-perform human traders in Continuous Double Auction (CDA) marketplaces similar to those used in the international financial markets. Cliff [4,5] used a GA to explore a continuous space of ZIP-trader auction-market mechanisms, where the space of possible auction-types explored included the CDA and also two purely one-sided mechanisms. Surprisingly, the GA would sometimes settle on novel *hybrid* auction mechanisms partway between the CDA and a one-sided auction. Such results occurred when the market's supply and demand schedules were unchanging, and also when the schedules undergo a single sudden "shock" change halfway through the evaluation process. These results could *prima facie* support the hypothesis that hybrid auctions are in general preferable to the well-known CDA mechanism. In this paper we present new results that clarify that hypothesis. In our new experiments, more than one shock-change in supply/demand occurs during the mechanism-evaluation process; under this regimen, CDA auction-mechanisms are identified by the GA as optimal in three of the four experiments reported here, and in the fourth experiment a near-CDA hybrid mechanism was evolved. From these results we conclude that, while evolved hybrid market-mechanisms may be useful in niches where the marketplace's supply and demand dynamics are known *a priori* to be relatively stable or regular, the CDA remains the mechanism of choice when it is difficult or impossible to make accurate predictions concerning the dynamic stability of the supply and demand schedules.

Keywords: Algorithmic Mechanism Design; Auction and Negotiation Technology; Automated Trading; ZIP Traders; Genetic Algorithms; e-Marketplaces.

Submitted in abridged form to the 2003 ACM Conference on Electronic Commerce (EC03), June 2003.

1. INTRODUCTION

ZIP (Zero-Intelligence-Plus) artificial trading agents, introduced in 1997 [1], are software agents (or "robots") that use simple machine learning techniques to adapt to operating as buyers or sellers in online open-outcry auction-market environments similar to those used in the seminal experimental economics work of Smith [16]. ZIP traders were originally developed as a solution to the pathological failures of Gode & Sunder's "ZI" (Zero-Intelligence) traders [10], but recent work by Das *et al.* at IBM [8] has shown that ZIP traders (unlike ZI traders) consistently out-perform human traders in human-against-robot experimental economics online e-marketplaces.

The operation of ZIP traders has been successfully demonstrated in experimental versions of continuous double auction (CDA) markets similar to those found in the international markets for commodities, equities, capital, and derivatives; and in posted-offer auction markets similar to those seen in domestic high-street retail outlets [1]. In any such market, there are a number of parameters that govern the adaptation and trading processes of the ZIP traders. In the original 1997 version of ZIP traders, the values of these parameters were set by hand, using "educated guesses". However, Cliff subsequently [2,3] presented the first results from using a standard genetic algorithm (GA) to automatically optimise these parameter values, thereby eliminating the need for skilled human input in deciding the values.

Prior to the research described in [4], in all previous work using artificial trading agents, ZIP or otherwise, the market mechanism (i.e., the type of auction the agents are interacting within) had been fixed in advance. Well-known market mechanisms from human economic affairs include: the English auction (where sellers stay silent and buyers quote increasing bid-prices), the Dutch Flower auction (where buyers stay silent and sellers quote decreasing offer-prices); the Vickery or second-price sealed-bid auction (where sealed bids are submitted by buyers, and the highest bidder is allowed to buy, but at the price of the *second-highest* bid); and the CDA (where sellers announce decreasing offer prices while *simultaneously and asynchronously* the buyers announce increasing bid prices, with the sellers being free to accept any buyer's bid at any time and the buyers being free to accept any seller's offer at any time). The CDA is of particular interest because it is the basis of most major national and international financial markets, and hence has been the subject of much academic study (see e.g., [9]).

Cliff [4] presented the first results from experiments where a genetic algorithm (GA) optimises not only the parameter values for the trading agents, but also the style of market mechanism in which those traders operate. To do this, a space of possible market mechanisms was created for evolutionary exploration. The space includes the CDA and also one-sided auctions similar (but not actually identical to) the English Auction (EA) and the Dutch Flower Auction (DFA). Significantly, this space is *continuously variable*, allowing for any of an *infinite* number of peculiar hybrids of these auction types to be evolved, which have no known correlate in naturally occurring (i.e., human-designed) market mechanisms. While there is nothing to prevent the GA from settling on solutions that correspond to the known CDA auction type or the EA-like and DFA-like one-sided mechanisms, it was found that hybrid solutions could lead to the most desirable market dynamics. Although the hybrid market mechanisms can easily be implemented in online electronic marketplaces, they have not been designed by humans: rather they are the product of evolutionary search through a continuous space of possible auction-types. Thus, the results in [4] were the first demonstration that radically new market mechanisms for artificial traders may be designed by automatic means.

This is not a trivial academic point: although the efficiency of the evolved market mechanisms are typically only a few percentage points better than those of the established human-designed mechanisms, the economic consequences could be highly significant. According to figures released by the New York Stock Exchange (NYSE), the total value of trades on the CDA-based NYSE for the year 2000 was $11060bn (i.e., a little over 11 trillion dollars). If only 0.1% of that liquidity could be eliminated or captured by a more efficient evolved market mechanism, the value saved (or profit generated) would still be in excess of $10bn. But that is just for one market: similar savings could be made at NASDAQ, at European exchanges such as LSE and LIFFE, and at similar exchanges elsewhere around the globe.

The key results in [4] have since been replicated [13,14] and qualitatively similar results have been demonstrated in non-ZIP trader e-marketplaces [17,6]. Results from a similar research project, using another evolutionary algorithm (i.e., genetic programming) for mechanism design in a different context, have also recently been published [12].

Section 2 gives a background overview of ZIP traders and of the experimental methods used, including a description of the continuously-variable space of auction types; it gives a brief review of our previous results and concludes with discussion of the rationale for the new experiments whose results are presented for the first time in this paper. Much of Section 2 is largely identical to the accounts given in previous papers, but is necessary here for completeness and comprehensibility. Our new results are presented in Section 3 and are discussed in Section 4.

2. BACKGROUND

2.1 Zero-Intelligence Plus (ZIP) Traders

ZIP traders are described fully in [1], which includes sample source-code in the *C* programming language. For the purposes of this paper a high-level description of the key parameters is sufficient. Each ZIP trader *i* is given a private (secret) limit-price, λ_i, which for a seller is the price below which it must not sell and for a buyer is the price above which it must not buy. If a ZIP trader completes a transaction at its λ_i price then it generates zero utility ("profit" for the sellers or "saving" for the buyers). For this reason, each ZIP trader *i* maintains a time-varying margin $\mu_i(t)$ and generates quote-prices $p_i(t)$ at time *t* according to $p_i(t)=\lambda_i(1+\mu_i(t))$ for sellers and $p_i(t)=\lambda_i(1-\mu_i(t))$ for buyers. The "aim" of traders is to maximise their utility over all trades, where utility is the difference between the accepted quote-price and the trader's λ_i value. Trader *i* is given an initial value $\mu_i(0)$ (i.e., $\mu_i(t)$ for $t=0$) which is subsequently adapted over time using a simple machine learning technique known as *the Widrow-Hoff rule* which is also used in back-propagation neural networks. This rule has a "learning rate" parameter β_i that governs the speed of convergence between trader *i*'s quoted price $p_i(t)$ and the trader's idealised "target" price $\tau_i(t)$. When calculating $\tau_i(t)$, traders introduce a small random absolute perturbation generated from $\boldsymbol{U}[0,c_a]$,[1] and also a small random relative perturbation coefficient generated from $\boldsymbol{U}[1-c_r,1]$ (when a trader is reducing its $p_i(t)$) or $\boldsymbol{U}[1,1+c_r]$ (when increasing $p_i(t)$) where c_a and c_r are global system constants. To smooth over noise in the learning system, there is an additional "momentum" parameter γ_i for each trader (such momentum terms are also commonly used in back-propagation neural networks).

Thus, adaptation in each ZIP trader *i* has the following parameters: initial margin $\mu_i(0)$; learning rate β_i; and momentum term γ_i. In an entire market populated by ZIP traders, these three parameters are assigned to each trader from uniform random distributions each of which is defined via "minimum" and "delta" values in the following fashion: $\mu_i(0)=\boldsymbol{U}(\mu_{min}, \mu_{min}+\mu_\Delta)$; $\beta_i=\boldsymbol{U}(\beta_{min}, \beta_{min}+\beta_\Delta)$; and $\gamma_i=\boldsymbol{U}(\gamma_{min}, \gamma_{min}+\gamma_\Delta)$; $\forall i$.

Consequently, to initialise an entire ZIP-trader market it is necessary to specify values for the six market-initialisation parameters μ_{min}, μ_Δ, β_{min}, β_Δ, γ_{min}, and γ_Δ; and also for the two system constants c_a and c_r. And so it can be seen that any set of initialisation parameters for a ZIP-trader market exists within an eight-dimensional real space, conventionally denoted by $\mathbf{R}^8$. Vectors in this 8-space can be considered as genotypes, and from an initial population of such genotypes it is possible to allow a GA to find new genotypes that best satisfy an appropriate evaluation function. This is exactly the process that was introduced in [2], as described in Section 2.3 below. Before that, we discuss the issue of simulating the passage of time.

The same discrete time-slicing approach was used here as was used in previous ZIP work [1,2,3,4,5,6,7]. In each time-slice, the atomic "significant event" is one quote being issued by one trader and the other traders then responding either by ignoring the quote or by one of the traders accepting the quote. (NB [8] used a continuous-time formulation of the ZIP-trader algorithm).

In the markets described here and in [1,2,3,4,5,6,7,13,14], on each time-slice a ZIP trader *i* is chosen at random from those currently able to quote (i.e. those who hold appropriate stock or currency), and trader *i*'s quote price $p_i(t)$ then becomes the "current quote" $q(t)$ for time *t*. Next, all traders *j* on the contraside (i.e. all buyers *j* if *i* is a seller, or all sellers *j* if *i* is a buyer) compare $q(t)$ to their own current quote price $p_j(t)$ and if the quotes cross (i.e. if

[1] Note that in this paper $v=\boldsymbol{U}[x,y]$ denotes a random real value *v* generated from a uniform distribution over the range *[x,y]*.

$p_j(t)<=q(t)$ for sellers, or if $p_j(t)>=q(t)$ for buyers) then the trader j is able to accept the quote. If more than one trader is able to accept, one is chosen equiprobably at random to make the transaction. If no traders are able to accept, the quote is regarded as "ignored". Once the trade is either accepted or ignored, the traders update their $\mu(t)$ values using the learning algorithm outlined above, and the current time-slice ends. This process repeats for each time-slice in a trading period, with occasional injections of fresh currency and stock, or redistribution of λ_i limit prices, until either a maximum number of transactions have occurred, or until either no seller or no buyer is able to quote, or until a maximum number of time-slices have passed since the last accepted quote (i.e., a until a protracted sequence – typically 100 – of successive ignored quotes occurs).

2.2 A Space of Possible Auctions

Now consider the case where we implement a ZIP-trader CDA market. In any one time-slice in a CDA either a buyer or a seller may quote, and in the definition of a CDA a quote is equally likely from each side. One way of implementing a CDA is, at the start of each time-slice, to generate a 50/50 random binary variable to determine whether the quote will come from a buyer or a seller, and then to randomly choose one individual as the quoter from whichever side the binary value points to.

Let $Q=b$ denote the event that a buyer quotes on any one time-slice and let $Q=s$ denote the event that a seller quotes, then for the CDA we can write $Pr(Q=s)=0.5$ and note that because $Pr(Q=b)=1.0-Pr(Q=s)$ it is only necessary to specify $Pr(Q=s)$, which we will abbreviate to Q_s hereafter. Note additionally that in an EA we have $Q_s=0.0$, and in the DFA we have $Q_s=1.0$. Thus, there are at least three values of Q_s (i.e. *0.0, 0.5,* and *1.0*) that correspond to three types of auction familiar from centuries of human economic affairs.

However, although the ZIP-trader case of $Q_s=0.5$ is indeed a good approximation to the CDA, the fact that any ZIP trader j will accept a quote whenever $q(t)$ and $p_j(t)$ cross means that the one-sided extreme cases $Q_s=0.0$ and $Q_s=1.0$ are not exact analogues of the EA and DFA. Nevertheless, consider the implications of considering values of Q_s of 0.0, 0.5, and 1.0 not as three distinct market mechanisms, but rather as three points on a *continuum*. How do we interpret, for example, $Q_s=0.1$? Certainly there is a straightforward implementation: on the average, for every nine quotes by buyers, there will be one quote from a seller. Yet the history of human economic affairs offers no examples (as far as we are aware) of such markets: why would anyone suggest such a bizarre way of operating, and who would go to the trouble of arbitrating (i.e., acting as an auctioneer for) such a mechanism? Nevertheless, there is no *a priori* reason to argue that the three known points on this Q_s continuum are the only loci of useful auction types. Maybe there are circumstances in which values such as $Q_s=0.1$ are preferred. Given the infinite nature of a real continuum, it seems appealing to use an automatic exploration process, such as a GA, to identify useful Q_s values.

Thus, a ninth dimension was added to the search space, and the genotype in the GA is now the eight real values for ZIP-trader initialisation, plus a real value for Q_s, so the GA is searching for points in $\mathbf{R}^9$ that give the best market dynamics.

2.3 The Genetic Algorithm

A simple genetic algorithm was used (see e.g. [11] for an introduction to genetic algorithms). As with each experiment reported in [2,3,4,5,6,7] a population of size 30 was used, and evolution was allowed to progress for some number of generations n_g. In each generation, all individuals were evaluated and assigned a "fitness" value (reflecting how good that genotype's market dynamics were); and the next generation's population was then generated via mutation and crossover on parents identified using rank-based tournament selection. Elitism (where an unaltered copy of the fittest individual from generation g is inserted into the population of $g+1$) was also used.

The genome of each individual was simply a vector of nine real values. In each experiment, the initial random population was created by generating random values from $\boldsymbol{U}[0,1]$ for each locus on each individual's genotype. Crossover points were between the real values, and crossover was governed by a Poisson random process with an average of between one and two crosses per reproduction event. Mutation was implemented by adding random values from $\boldsymbol{U}[-m(g),+m(g)]$ where $m(g)$ is the mutation limit at generation g (starting the count at $g=0$). Mutation was applied to each locus in each genotype on each individual generated from a reproduction event, but the mutation limit $m(g)$ was gradually reduced via an exponential-decay annealing function of the form: $log_{10}(m(g))=-(log_{10}(m_s)-(g/(N_g-1))log_{10}(m_s/m_e))$ where N_g is the maximum number of generations and m_s is the "start" mutation limit (i.e., for $m(0)$) and m_e is the "end" mutation limit (i.e., for $m(n_g-1)$). In all the experiments reported here and in earlier papers [7,8], $N_g=10^3$, $m_s=0.05$, and $m_e=0.0005$.

If ever mutation caused the value at a locus to fall outside $[0.0,1.0]$ it was simply clipped to stay within that range. This clip-to-fit approach to dealing with out-of-range mutations biases evolution toward extreme values (i.e. the upper and lower bounds of the clipping), and so Q_s values of 0.0 or 1.0 are, if anything, more likely to be evolved. Initial and mutated genome values of μ_Δ, β_Δ, and γ_Δ were also clipped to satisfy $(\mu_{min}+\mu_\Delta)<=1.0$, $(\beta_{min}+\beta_\Delta)<1.0$, & $(\gamma_{min}+\gamma_\Delta)<1.0$.

The fitness of genotypes was evaluated using the same methods as described in [2,3,4,5,6,7]. One *trial* of a particular genome was performed by initialising a ZIP-trader market from the genome, and then allowing the traders to operate within the market for a fixed number of trading periods, with allocations of stock and currency being replenished between trading periods. Each trading period was ended in the manner described above.

During each trading period, Smith's α measure [16] of deviation of transaction prices from the theoretical market equilibrium price was monitored, and a weighted average was calculated across the trading periods in the trial. As the outcome of any one such trial is influenced by stochasticity in the system, the final fitness value for an individual was calculated as the arithmetic mean of 100 such trials. Note that as minimal deviation of transaction prices from the theoretical equilibrium price is desirable, lower scores are better: the intention here is to *minimise* the fitness value.

In [4] the number of generations n_g for each experiment was set to equal N_g (i.e., 1000), but all the significant evolutionary activity was found to occur in the first 500 generations; hence n_g=500 was used in subsequent work [5,6,7] and here also. Thus, in any one experiment, there are 30 individuals evaluated over 500 generations where each evaluation involves calculating the mean of 100 trials, so a total of 1.5 million market trials would be executed in any one GA experiment. Nevertheless, the progress of each GA experiment is itself affected by stochasticity (e.g. the GA may become trapped on local optima) and so to generate reliable results each experiment was repeated 50 times, requiring a total of 75 million market trials. On a current single-CPU PC, 50 repetitions of the multi-shock experiments described in Section 3 take around twelve days to complete, so the results presented in Section 3 would have required a total of 96 days of continuous computation time had a single-CPU PC been used. As shown by Walia [17], much of this consumption of compute time is due to our choice of a computationally expensive (but statistically rigorous) random number generator function.

2.4 Previous Results

In our first evolving-mechanism paper [4], three differing market supply and demand schedules were used, referred to as markets M1, M2, and M3 respectively. Figures 1 and 2 show markets M1 and M2. Market M3 is very similar to M1, having the same equilibrium point and the same number of traders; the only difference being that in M1 the difference between successive limit prices for the traders is 0.25, while in M3 it is 0.30: see [4].

Figures 1 and 2 both show a supply and demand schedule for a marketplace with 11 buyers and 11 sellers, each empowered to buy/sell one unit of commodity, and both are similar (or identical) to the schedules used by Smith [16]. Figure 3 shows results from 50 repetitions of an experiment where the GA explores the $\mathbf{R}^9$ subspace in an attempt to optimize the ZIP-trader market parameters for operating in M1: for each experiment, the fitness of the best (elite) member of the population is recorded.

The results are clearly tri-modal. Of the 50 repetitions, in five the elite ends up on fitness minima of about 3.2, while the other two elite fitness modes are on less-good minima of around 4.0 and 4.75. For comparison, Figure 5 shows the results of 50 repeats of the same experiment, where the value of Q_s was *not* evolved, being instead clamped at 0.5: i.e. the CDA value. The CDA mechanism is often applauded as an auction mechanism in which equilibration is rapid and stable, so we could expect the best fitness from using this market type. With the fixed CDA auction style, an average elite fitness of around 4.5 is settled on by the majority of experiments (48 repetitions) while a small minority (2 repetitions) settle on a less good mode of around 5.1. Clearly then, the evolved-mechanism results are better than the fixed-mechanism CDA results; that is, when the GA is allowed to find its own value of Q_s rather than have the CDA Q_s value of 0.5 imposed on it, it finds fitter solutions – solutions with less deviation of transaction prices from the equilibrium price. As it happens, the Q_s value found in the best elite mode for the evolving-mechanism M1 experiments is zero [4], and for M2 the best Q_s was also zero [4]. But, surprisingly and significantly, for M3 the best Q_s was neither zero, nor 0.5, nor 1.0 – i.e. none of the Q_s values corresponding to traditional human-designed auction mechanisms; rather, the best Q_s for M3 was found to be around 0.16 [4]. The fact that schedules M1 and M3 are superficially very similar, while their optimal Q_s values are significantly different, is explored in depth in [7], where detailed visualizations of the underlying GA fitness landscapes are shown.

All of the results in [4] came from experiments in which the same static supply/demand schedule was used for the duration of each evaluation of every genotype. This is a somewhat unrealistic simplification, for two reasons. First, a primary reason why auction mechanisms such as the CDA are of interest is their ability to adapt to *changes* in the market's supply and demand curves. Second, it is likely that the GA exploited this regularity and *over-fitted* the ZIP-trader parameters to the particular market schedules used (e.g. a genome that does well in M1 might perform poorly in M2). Thus, in a subsequent paper [5], similar experiments were run but in these new studies the evaluation of a genotype involved six trading periods on one market schedule, followed by a shock-change to another schedule, and then another six trading periods on the new schedule; with the fitness of the genotype being calculated over the entire 12 periods of trading.

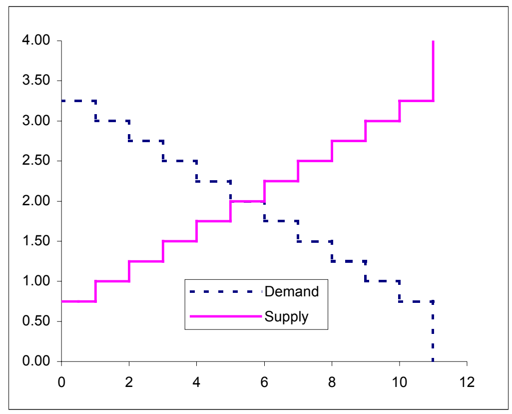

Figure 1: Supply and demand schedules for market M1. Vertical axis is Price; horizontal axis is Quantity.

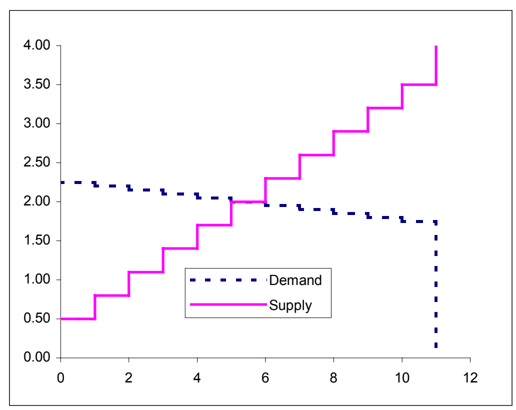

Figure 2: Supply and demand schedules for market M2.

Hence in these experiments the genotypes had to optimize not only the ZIP-trader's *ab initio* adaptation to the first schedule but also their re-adaptation to the new schedule introduced half-way through the evaluation process. Two sets of experiments were performed: one in which the ZIP traders operated in M1 for the first six periods followed by a shock-change to M2 for the final six periods (referred to as the M1M2 experiments); and another in which the order was M2 followed by M1 (referred to as M2M1). The order was significant: the M1M2 results differed significantly from the M2M1 results. Although in the single-schedule experiments both M1 and M2 were found to have optima at Q_s=0, when the two schedules were both used in one trial then non-zero values of Q_s evolved: for M1M2 the best-mode value was a "hybrid" of around 0.25; while for M2M1 the best value was 0.45, which did not yield statistically significant differences in performance from the CDA value of 0.5. For full details, see [5].

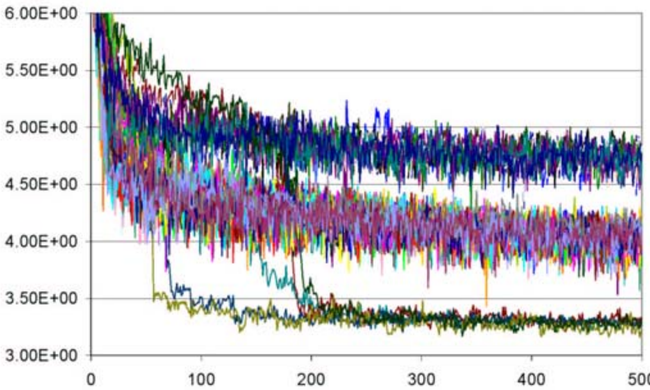

Figure 3: Elite fitness values from 50 repetitions of a 500-generation evolving-mechanism (EM) experiment operating with M1. Horizontal axis is generation-number; vertical axis is fitness. Lower fitness values are better solutions (less deviation from equilibrium). Results are tri-modal, with five of the repetitions (10%) settling to values around 3.2.

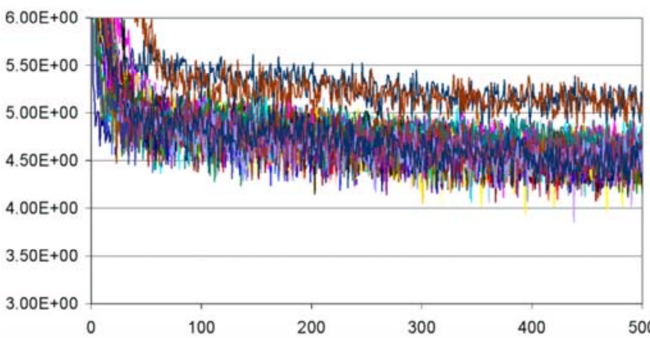

Figure 4: Elite fitness values from 50 repetitions of a 500-generation experiment operating with M1, but with a fixed-mechanism (FM) CDA of Q_s=*0.5*: bimodal results, with 96% of the repetitions settling to fitness values around 4.5 and the remaining 4% at around 5.2.

2.5 Rationale for multi-shock evaluation

The M1M2 and M2M1 results in [5] indicate that the genotype (market mechanism and trader parameters) best suited to dealing with a single shock-change from M1 to M2 differs from the genotype best suited to dealing with the single shock-change from M2 to M1. These results raise the possibility that the GA had over-fitted the genotypes to the particular single market shock imposed during the evaluation process. To explore this possibility, new experiments were conducted where multiple shocks occur during evaluation. For ease of comparison to the previous results, we continue to work with M1, M2, and M3.

Results from four sets of experiments are reported in the next section: one set is referred to as "M121" to denote the fact that market M1 was used for the first six periods followed by M2 for the second six periods followed by M1 again for the final six periods. The second set of experiments, where six periods of M2 were followed by six periods of M1 followed by a final six periods of M2, is referred to as M212. It should be noted that in both the M121 experiments and the M212 experiments the genotype has to be adapted to *both* the transition from M1 to M2 *and* the transition from M2 to M1. Having established in [7] that M1 and M3 can yield different evolved solutions despite their very close similarity, two further sets of experiments were conducted in which M3 was substituted for M1. These are referred to as the M123 and M321 experiments.

In [2,3,4] the evaluation function was a weighted average of Smith's α measure of root mean square deviation of transaction prices from the theoretical equilibrium price: in each trading period p the value α_p was calculated, and the fitness score was computed as $(1/w_s).\Sigma(\alpha_p.w_p)$ for p=1…6 with weights w_1=1.75, w_2=1.5, w_3=1.25, and w_4=w_5=w_6=1.0; and w_s=Σw_p. In the M1M2 and M2M1 dual-schedule experiments reported in [5], this was simply extended so that p=1…12 and $w_{p>6}$=w_{p-6}. Similarly, in the triple-schedule experiments reported here, the extension was simply that p=1…18 and $w_{p>12}$=w_{p-12}.

3. MULTI-SHOCK RESULTS

Figure 5 shows results from 50 repetitions of the M121 evolving-mechanism (EM) experiment: as before, for each repetition the fitness of the best (elite) member of the population is recorded at each generation. The results are bimodal: At g=500 the majority (n=48) mode has mean fitness of 4.41 and a standard deviation (s.d.) of 0.265. Figure 6 shows the evolutionary trajectory of the elite genotype's value of Q_s at each generation: clearly, the best elite mode uses an auction mechanism indistinguishable from the Q_s=0.5 CDA. When the M121 experiments are re-run with a fixed-mechanism (FM) of Q_s=0.5, at g=500 the best (n=49) mode has a mean fitness of 4.33 and a s.d. of 0.786.

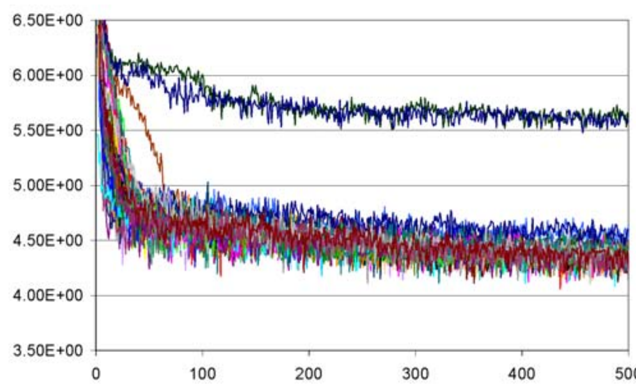

Figure 5: Elite fitness values from 50 repetitions of the 500-generation M121 EM experiment. Format as for Figure 3.

Figure 7 shows results from 50 repetitions of the M212 EM experiment. The results are unimodal. At g=500 the mean fitness is 3.92 with s.d.=0.0734. Figure 8 shows the evolutionary trajectory of the elite-genotype Q_s values: clearly, the best elite mode again uses an auction mechanism indistinguishable from the

Q_s=0.5 CDA. When the M212 experiments are re-run with a fixed-market (FM) of Q_s=0.5 the best (n=49) mode has a mean fitness of 3.83 and an s.d. of 0.109 at g=500.

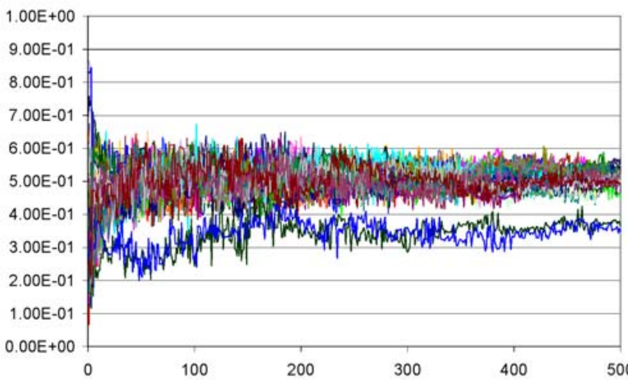

Figure 6: Evolutionary trajectory of elite Q_s values from the 50 M121 EM experiments shown in Fig. 5. Horizontal axis is gen-eration number; vertical axis is elite-genotype Q_s value. At g=500 the mean Q_s in the n=48 best elite mode is 0.509 with s.d.=0.0227.

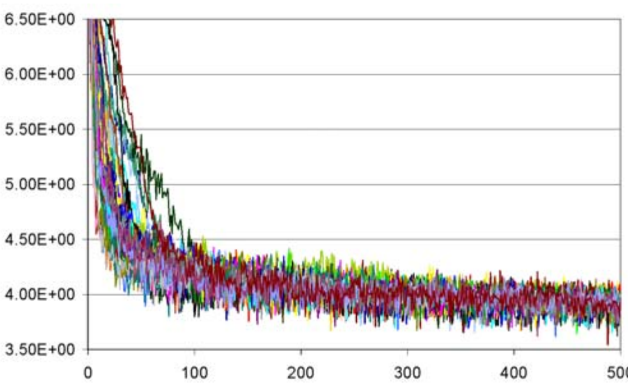

Figure 7: Elite fitness values from n=50 repetitions of the 500-generation M212 EM experiment. Format as for Figure 3.

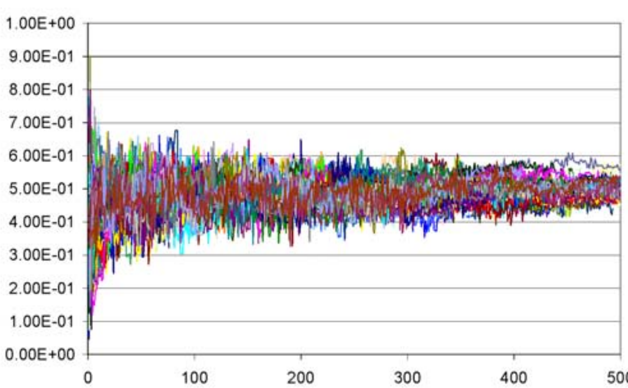

Figure 8: Evolutionary trajectory of Q_s values from the 50 M212 EM experiments shown in Fig. 7. Format as for Fig. 6. At g=500 the mean Q_s in the n=50 best elite mode is 0.497, with s.d.=0.0263.

The Wilcoxon-Mann-Whitney test [15] was used to investigate the null hypothesis H_0 that the distribution of elite fitness values at generation 500 is not statistically different between the 50 values for the evolving-market (EM) case and the 50 values for the CDA fixed-marked (FM) case for both the M121 data and the M212 data; for both cases, the H_1 hypothesis was that the EM data were stochastically smaller than the FM data (i.e. that the evolving-mechanism genotypes gave better market dynamics).

For M121 the value of z was -1.9544, for which p=~0.0253 and so this is not significant at the 1% level, while for M212 the value of z was -0.0517, which is manifestly not significant. And so for both M121 and M212 we accept H_0: the elite fitness results from the EM experiments are not statistically different from the FM Q_s=0.5 data.

Figure 9 shows the elite fitness values from 50 M321 EM experiments, where at g=500 the best (and only) elite-mode has mean fitness of 3.98 and a s.d. of 0.0501. Figure 10 shows the corresponding elite-Q_s plot, where at g=500 the 50 elite genotype Q_s values have a mean of 0.473 and a s.d. of 0.0218. Again, these evolved values of Q_s are so close to the CDA value of Q_s=0.5 that it is no surprise to learn that the FM results were also uni-modal with the g=500 mean fitness also being 3.98 (s.d.=0.0673). For completeness, the Wilcoxon-Mann-Whitney z value was –0.169 which is again not significant at the 1% level, so there is no statistically significant difference between the final EM and FM results for M321.

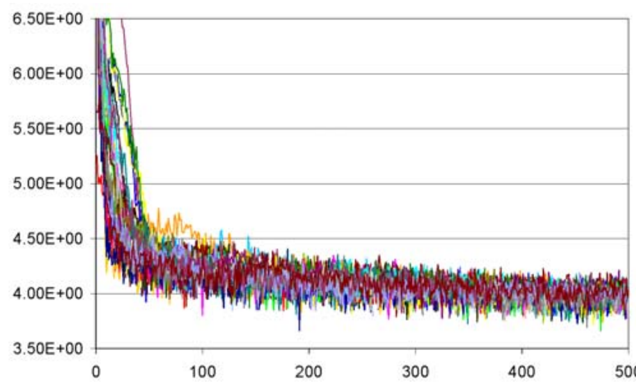

Figure 9: Elite fitness values from n=50 repetitions of the 500-generation M321 EM experiment. Format as for Figure 3.

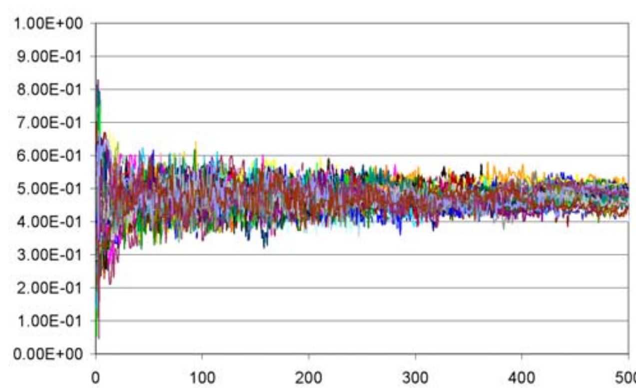

Figure 10: Evolutionary trajectory of Q_s values from the 50 M321 EM experiments shown in Fig.9. Format as for Fig.6.

Thus, the results for M121, M212, and for M321 all indicate that the CDA cannot be improved upon: when the market mechanism is placed under evolutionary control, the final evolved mechanism is statistically indistinguishable from the CDA. Curiously though, the results from M123 are qualitatively different.

Figure 11 shows the elite fitness values from 50 M123 EM experiments, where at g=500 the best (and only) elite-mode has mean fitness of 4.24 and a s.d. of 0.0664. Figure 12 shows the corresponding elite-Q_s plot, where at g=500 the 50 elite genotype Q_s values have a mean of 0.564 and a s.d. of 0.0238. Superficially these evolved values of Q_s appear to be close to the CDA value of Q_s=0.5, but it is worth noting that there are 2.5 standard deviations between 0.5 and 0.564. The M123 FM fitness results were also uni-modal with the g=500 mean being 4.28 (s.d.=0.0763). Intriguingly, the Wilcoxon-Mann-Whitney z value was –3.209, for which p=~0.0007, and that is clearly significant at the 1% level, so there *is* a statistically significant difference between the final EM and FM results for M123: that is, the evolving-mechanism results are statistically smaller (i.e., better) than the fixed-mechanism CDA results.

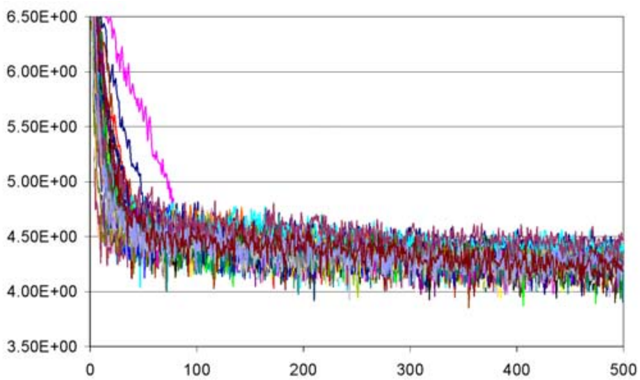

Figure 11: Elite fitness values from n=50 repetitions of the 500-generation M123 EM experiment. Format as for Fig. 3.

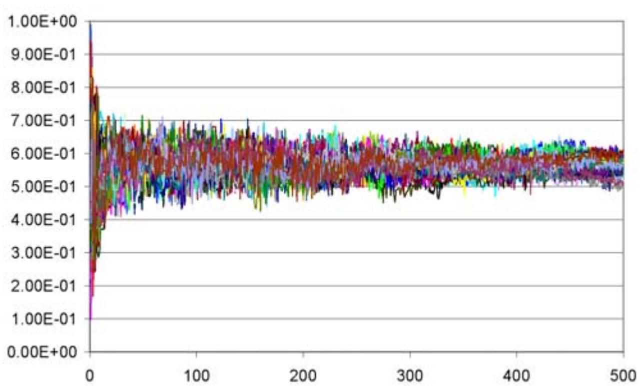

Figure 12: Evolutionary trajectory of Q_s values from the 50 M123 EM experiments shown in Fig.11. Format as for Fig. 6.

The nature of the difference between the M123 EM and FM results is made more clear in Figure 13, which shows a scatter plot of the fitness of the elite genotype at g=500 as a function of the elite genome's Q_s value for the 50 M123 EM experiments and the 50 M123 FM experiments. While the FM data are roughly symmetric about their mean fitness, the EM data has a skew distribution. By visual examination it is clear that in the fitness range [4.10,4.20], there are 6 FM data points and 14 EM data points; in the fitness range [4.20,4.30] there are 25 FM data points and 31 EM data points; in the range [4.30,4.40] there are 16 FM points and 3 EM points; while in the range [4.40,4.50] there are 3 FM data points and 2 EM data points.

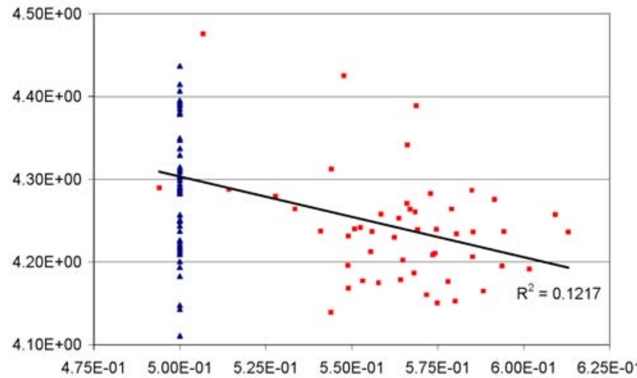

Figure 13: Scatter-plot of g=500 elite-genotype Q_s values (horizontal axis) against fitness (vertical axis) scores from the 50 M123 EM experiments (Fig.11) and the 50 M123 FM experiments (Fig.12). The 50 FM data are the vertical line at Q_s=0.50; the EM data are the cloud of points for which $Q_s \neq$0.50. The regression line shows a weak (r^2=0.122) negative correlation between the EM Q_s values and fitness scores.

4. DISCUSSION AND CONCLUSION

This paper is the third in a series that explores the use of evolutionary algorithms for the automated design of auction-market mechanisms. In [4] it was established that values of Q_s=0.0, corresponding to one-sided auctions, were optimal for M1 and M2 individually, and in [5] it was demonstrated that when the evolving ZIP traders have to deal with a single "shock" transition from M1 to M2, or from M2 to M1, two-sided mechanisms are found by the GA to give the most efficient markets: for M1M2 the hybrid Q_s=~0.25 market gives significantly better results than the CDA, while for M2M1 the evolved solution of Q_s=~0.45 was no better *but also no worse than* the CDA.

This paper has presented results from four new pairs of experiments (requiring 96 days of continuous CPU time) in which the GA is used to discover genotypes that are well-adapted to dealing with multiple shocks in the evaluation process. The new M121 and M212 results presented here have demonstrated that, when the evaluation process involves "shock" jumps from M1 to M2 *and* from M2 to M1, the CDA of Q_s=0.5 cannot be improved upon by the GA. The M321 results similarly indicated that the CDA is the mechanism that gives the best market dynamics in the presence of multiple shocks.

But the M123 results demonstrate that, even in the presence of multiple market shocks, hybrid auction mechanisms can be optimal. It is notable that in this case the evolved optimal value of Q_s differs from the CDA value of 0.50 by only a small amount: the mean elite Q_s value is 0.56, although Figure 13 indicates that the distribution is skew with respect to this mean, and the slight negative correlation between fitness and Q_s implies that the upper percentiles of the elite evolved genomes will have values of Q_s>0.56. The precise reason why such a small bias or deviation away from the CDA can give slightly better market dynamics remains a topic for further research.

Thus, this paper extends the line of research initiated in [4] by again demonstrating the use of an evolutionary search through an infinite space of possible market designs that includes the CDA of $Q_s=0.5$ and also the two pure one-sided solutions of $Q_s=0.0$ and $Q_s=1.0$. Again a new "hybrid" market mechanism was found to give better market dynamics than the previously-known auction styles. Once more, while such evolved market mechanisms are unlike any human-designed mechanism, they could nevertheless readily be implemented as online electronic marketplaces.

Viewing the results presented here in the context of the earlier no-shock and single-shock results presented in [4] and [5], it is notable that as more shocks are introduced into the evaluation process, so it appears that there is a narrowing of the difference between the GA-evolved optimal Q_s and the CDA value of Q_s=0.5. That is, as more shocks are introduced there is a *convergence* toward the CDA. It is tempting to hypothesize that, in the limit where *any* number/type of shocks can occur during evaluation/trading, the CDA is the optimal market mechanism. But a corollary to that hypothesis is the claim that if there are *known to be constraints* on the number or type of shocks that occur in the marketplace's supply and demand during trading, then the optimal market mechanism *may* be a non-CDA hybrid.

The major contribution of the M123 results presented here is to demonstrate that, even when there are *multiple* shock changes in supply and demand, there may be sufficient exploitable regularity in the supply and demand schedules of some online marketplaces that non-CDA hybrid auction mechanisms are more efficient than any human-designed market mechanism, including the CDA. Given these results, coupled with the demonstration by Das *et al.* [8] that ZIP trading agents reliably outperform human traders in experimental CDA settings, it seems plausible to conjecture that, in future, some or possibly all major financial markets will be implemented as e-marketplaces populated by autonomous software-agent traders. In such an agent-dominated future, market mechanisms originally designed by humans for human traders may not be the most efficient, and the results of this paper demonstrate that new hybrid mechanisms can be evolved that are more efficient than traditional human-designed markets.

Even if such hybrids are only a few percentage points more efficient than conventional human-designed mechanisms, it seems perfectly plausible that the results of using these artificially-evolved auction-mechanism designs (populated by artificial trading agents) in major international financial markets will be savings or profits measured in billions of dollars.

5. REFERENCES

[1] Cliff, D., "Minimal Intelligence Agents for Bargaining Behaviours in Market Environments" Technical Report HPL-97-91, Hewlett-Packard Laboratories, 1997.

[2] Cliff, D., "Genetic optimization of adaptive trading agents for double-auction markets" in *Proc. Computational Intelligence in Financial Eng. (CIFEr98).* IEEE/IAFE /Informs, pp.252-258, 1998.

[3] Cliff, D., "Evolutionary optimisation of parameter sets for adaptive software-agent traders in continuous double-auction markets", Technical Report HPL-2001-99, HP Labs, 2001.

[4] Cliff, D., "Evolution of market mechanism through a continuous space of auction types", Technical Report HPL-2001-326, Hewlett-Packard Laboratories, 2001. Presented at *Computational Intelligence in Financial Engineering (CIFEr02)* track of WCCI2002, Honolulu, May 2002.

[5] Cliff, D., "Evolution of market mechanism through a continuous space of auction-types II: Two-sided auction mechanisms evolve in response to market shocks", HP Labs Technical Report HPL-2002-128. Presented at *Agents for Business Automation* (ABA02), Las Vegas, June 2002.

[6] Cliff, D., Walia, V., & Byde, A. "Evolved Hybrid Auction Mechanisms in Non-ZIP Trader Marketplaces", Technical Report HPL-2002-247, Hewlett-Packard Laboratories, (Submitted to *CIFEr03*), 2002.

[7] Cliff D., "Visualising Search-Spaces for Evolved Hybrid Auction Mechanisms" Technical Report HPL-2002-291, Hewlett-Packard Laboratories. To be presented at the *Beyond Fitness: Visualising Evolution* workshop, at the 8th International Conference on the Simulation and Synthesis of Living Systems; Australia, Dec. 2002.

[8] Das, R., Hanson, J. E., Kephart, J. O., & Tesauro, G., "Agent Human Interaction in a Continuous Double Auction", In *Proceedings of the International Joint Conferences on Artificial Intelligence*. Seattle, USA, August 2001.

[9] Friedman, D., & Rust, J., (eds). *The Double Auction Market: Institution, Theories, and Evidence*. Addison Wesley, 1993.

[10] Gode, D., & Sunder, S., "Allocative efficiency of markets with zero-intelligence traders" *Journal of Political Economy* **101**:119-137, 1993.

[11] Mitchell, M., *An Introduction to Genetic Algorithms.* MIT Press, 1998.

[12] Phelps, S., McBurney, P., Parsons, S., & Sklar, E., "Co-evolutionary auction mechanism design: a Preliminary Report", Technical Report ULCS-02-019, Department of Computer Science, University of Liverpool, 2002.

[13] Qin, Z., *Evolving Marketplace Designs by Artificial Agents.* MSc Thesis, Computer Science Dept., Bristol Uni., 2002.

[14] Robinson, N., *Evolutionary Optimization of Market-Based Control Systems for Resource Allocation in Compute Farms.* MSc Thesis, University of Sussex School of Cognitive and Computing Sciences, Sep. 2002.

[15] Siegel, S. & Castellan, N. J., *Nonparametric Statistics for the Behavioural Sciences*. McGraw Hill, 1988.

[16] Smith, V., "Experimental study of competitive market behavior". *Journal of Political Economy* **70**:111-137, 1962

[17] Walia, V., *Evolving Market Design*, MSc Thesis, School of Computer Science, University of Birmingham, Sep. 2002.

6. APPENDIX

Eight figures excluded from Section 3 for reasons of brevity are presented here. Figures 14 to 18 show the evolutionary trajectories of the elite fitness values for the CDA Q_s=0.5 fixed-market (FM) experiments for M121, M212, M321, and M123 respectively. The format for all four is the same as for Figure 3.

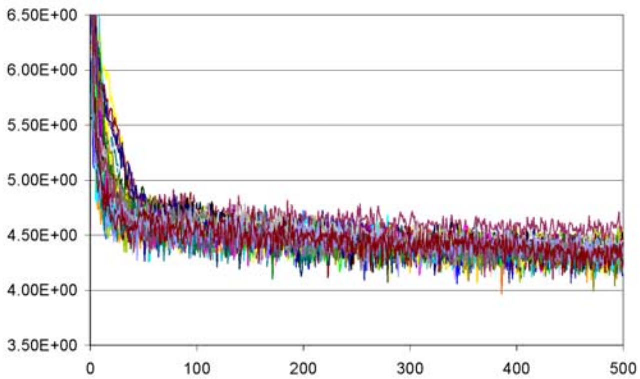

Figure 14: Elite fitness values from n=50 repetitions of the 500-generation M121 FM Q_s=0.5 experiment.

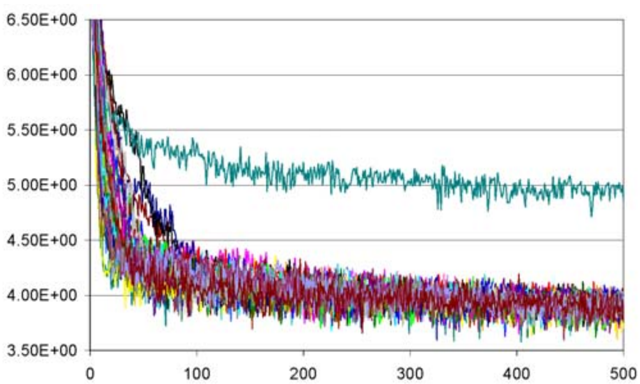

Figure 15: Elite fitness values from n=50 repetitions of the 500-generation M212 FM Q_s=0.5 experiment.

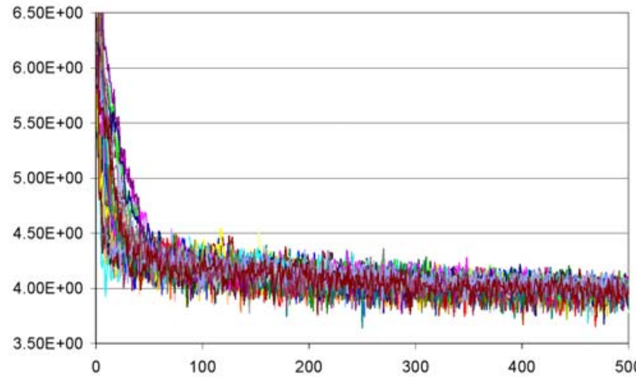

Figure 16: Elite fitness values from n=50 repetitions of the 500-generation M321 FM Q_s=0.5 experiment.

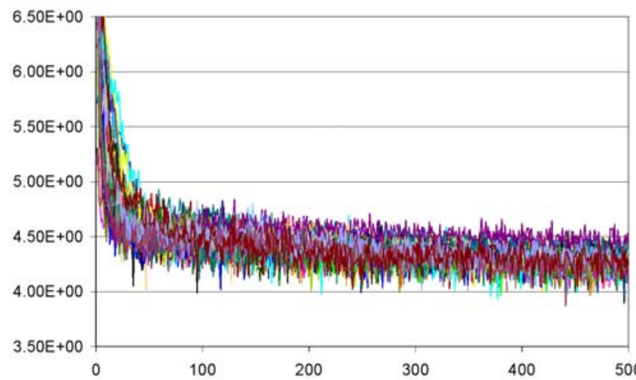

Figure 17: Elite fitness values from n=50 repetitions of the 500-generation M123 FM Q_s=0.5 experiment.

Figures 18, 19, 20 and 21 show the evolutionary trajectory of the mean (plus and minus one s.d.) Q_s value in the best elite fitness mode from the EM experiments for M121, M212, M321, and M123 respectively.

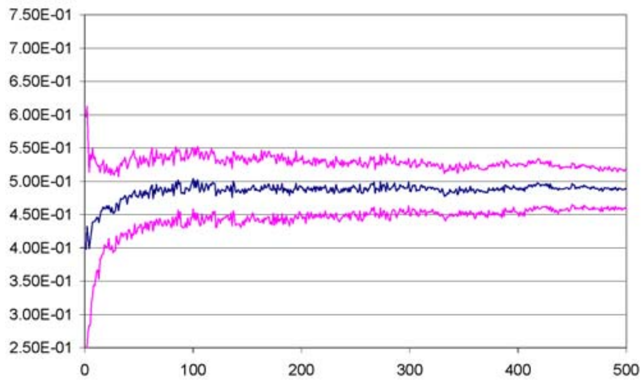

Figure 18: Evolutionary trajectory of the mean (plus and minus one s.d.) best elite-mode Q_s value in the M121 EM experiment (cf. Figure 6).

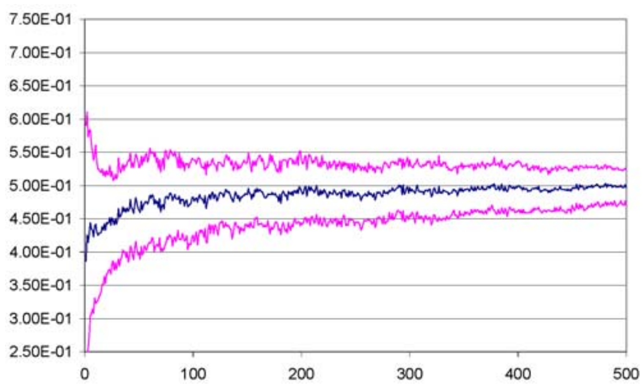

Figure 19: Evolutionary trajectory of the mean (plus and minus one s.d.) best elite-mode Q_s value in the M212 EM experiment (cf. Figure 8).

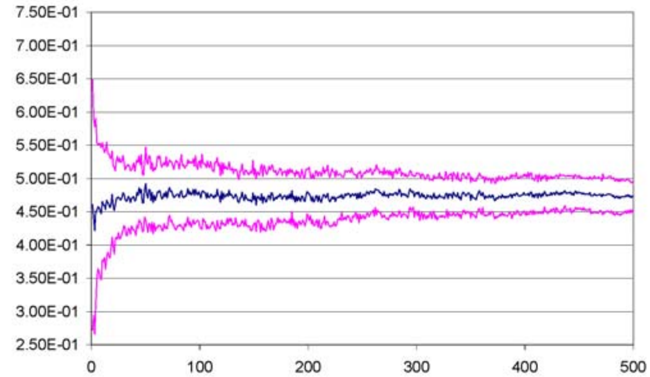

Figure 20: Evolutionary trajectory of the mean (plus and minus one s.d.) best elite-mode Q_s value in the M321 EM experiment (cf. Figure 10).

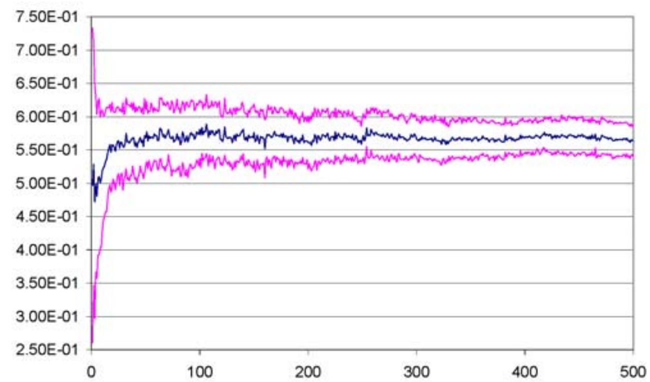

Figure 21: Evolutionary trajectory of the mean (plus and minus one s.d.) best elite-mode Q_s value in the M123 EM experiment (cf. Figure 12).