Evolution of market mechanism through a continuous space of auction-types II: Two-sided auction mechanisms evolve in response to market shocks

Dave Cliff
Information Infrastructure Laboratory
HP Laboratories Bristol
HPL-2002-128
May 8th , 2002*

E-mail: dave_cliff@hp.com

market mechanism, market design, ZIP traders, genetic algorithms, e-Marketplaces

This paper describes the use of a genetic algorithm (GA) to find parameter-values for trading agents that operate in virtual "e-marketplaces", where the rules of the marketplaces are also under simultaneous control of the GA. The aim is to use the GA to automatically design new agent-based e-marketplaces that are more efficient than markets designed by (or populated by) humans. Das *et al*. (2001) recently demonstrated that ZIP software-agent traders consistently outperform human traders in Continuous Double Auction (CDA) marketplaces. Cliff (2001b) used a GA to explore a continuous space of auction mechanisms, with ZIP traders simultaneously evolving to operate efficiently in these evolved markets. The space of possible auction-types explored includes the CDA and also two purely one-sided mechanisms. Surprisingly, the GA did not settle on the CDA. Instead, in two experiments, optima were found at a one-sided auction mechanism; and in a third experiment a novel hybrid auction mechanism partway between the CDA and a one-sided auction was evolved. This paper extends that research by studying the auction mechanisms that evolve when the market supply and demand schedules undergo a sudden "shock" change half-way through the evaluation process. It is shown that hybrid market mechanisms (again partway between the CDA and a one-sided mechanism) can evolve in place of the one-sided solutions that evolve when there are no market shocks. Furthermore it is demonstrated that the precise nature of the hybrid auction that evolves is dependent on the nature of the shock. These results indicate that the evolution of one-sided mechanisms reported by Cliff (2001b) is an artefact of using single fixed schedules, and that in general two-sided auctions will evolve. These two-sided auctions may be hybrids unlike any human-designed auction and yet may also be significantly more efficient than any human designed market mechanism.

* Internal Accession Date Only
Approved for External Publication
ABA 02: Agents for Business Automation, Las Vegas, June '02

Evolution of market mechanism through a continuous space of auction-types II: Two-sided auction mechanisms evolve in response to market shocks

Dave Cliff

dave_cliff@hp.com

Hewlett-Packard Laboratories Bristol, Filton Road, Bristol BS34 8QZ, England, U.K.

Abstract: This paper describes the use of a genetic algorithm (GA) to find parameter-values for trading agents that operate in virtual "e-marketplaces", where the rules of the marketplaces are also under simultaneous control of the GA. The aim is to use the GA to automatically design new agent-based e-marketplaces that are more efficient than markets designed by (or populated by) humans. Das *et al.* (2001) recently demonstrated that ZIP software-agent traders consistently outperform human traders in Continuous Double Auction (CDA) marketplaces. Cliff (2001b) used a GA to explore a continuous space of auction mechanisms, with ZIP traders simultaneously evolving to operate efficiently in these evolved markets. The space of possible auction-types explored includes the CDA and also two purely one-sided mechanisms. Surprisingly, the GA did not settle on the CDA. Instead, in two experiments, optima were found at a one-sided auction mechanism; and in a third experiment a novel *hybrid* auction mechanism partway between the CDA and a one-sided auction was evolved. This paper extends that research by studying the auction mechanisms that evolve when the market supply and demand schedules undergo a sudden "shock" change half-way through the evaluation process. It is shown that hybrid market mechanisms (again partway between the CDA and a one-sided mechanism) can evolve in place of the one-sided solutions that evolve when there are no market shocks. Furthermore it is demonstrated that the precise nature of the hybrid auction that evolves is dependent on the nature of the shock. These results indicate that the evolution of one-sided mechanisms reported by Cliff (2001b) is an artefact of using single fixed schedules, and that in general two-sided auctions will evolve. These two-sided auctions may be hybrids unlike any human-designed auction and yet may also be significantly more efficient than any human designed market mechanism.

Keywords: Market Mechanism; Market Design; ZIP traders; Genetic Algorithms; e-Marketplaces.

Abridged version of this paper to be presented at *Agents for Business Applications* (ABA02) Las Vegas, June 2002.

I. INTRODUCTION

ZIP (Zero-Intelligence-Plus) artificial trading agents, introduced by (1997), are software agents (or "robots") that use simple machine learning techniques to adapt to operating as buyers or sellers in open-outcry auction-market environments similar to those used in the experimental economics work of Smith (1962). ZIP traders were originally developed as a solution to the pathological failures of Gode & Sunder's (1993) "ZI" (Zero-Intelligence) traders, but recent work by Das *et al.* (2001) at IBM has shown that ZIP traders (unlike ZI traders) consistently out-perform human traders in human-against-robot experimental economics 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 (Cliff, 1997). 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 (1998; 2001a) 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 by Cliff (2001b), 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: game-theoretic analysis demonstrates that this mechanism encourages honesty and is robust to attack by dishonest means); 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., Friedman & Rust, 1993).

Cliff (2001b) 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 can lead to the most desirable market dynamics. Although the hybrid market mechanisms could 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 Cliff (2001b) 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: see NYSE, 2002). 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. And that is just for one market: similar savings could presumably made at NASDAQ, at European exchanges such as LSE and LIFFE, and at similar exchanges elsewhere around the globe.

Section II gives an overview of ZIP traders and of the experimental methods used, including a description of the continuously-variable space of auction types. This is largely identical to the account given by Cliff (2001b), albeit extended to summarise the results from that paper and to describe how the new experiments whose results are presented here differ from the previous work. These new results are presented in Section III and are discussed in Section IV. Note that in this paper $v=\boldsymbol{U}[x,y]$ denotes a random real value v generated from a uniform distribution over the range $[x,y]$.

II. METHODS

A. Zero-Intelligence Plus (ZIP) Traders

ZIP trading agents were described fully in a lengthy report by Cliff (1997), which included 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 (i.e., 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 $\boldsymbol{U}[0,c_a]$ (this perturbation is positive for sellers, negative for buyers) and also a small random relative perturbation generated from $\boldsymbol{U}[1-c_r,1]$ (buyers) or $\boldsymbol{U}[1,1+c_r]$ (sellers). Here 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, values for these three parameters are randomly assigned to each trader via: $\mu_i(0)=\boldsymbol{U}(\mu_{min},\ \mu_{min}+\mu_\Delta)$; $\beta_i=\boldsymbol{U}(\beta_{min},\ \beta_{min}+\beta_\Delta)$; and $\gamma_i=\boldsymbol{U}(\gamma_{min},\ \gamma_{min}+\gamma_\Delta)$. Hence, to initialise an entire ZIP-trader market it is necessary to specify values for the six market-initialisation parameters $\mu_{min},\ \mu_\Delta,\ \beta_{min},\ \beta_\Delta,\ \gamma_{min},$ and γ_Δ; and also for the two global 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. 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 genotype vectors that best satisfy an appropriate evaluation function. This is exactly the process that was introduced by Cliff (1998, 2001a), and that is described further below.

When monitoring events in a real auction, as more precision is used to record the time of events, so the likelihood of any two events occurring at exactly the same time is diminished. For example, if two bid-quotes made at five minutes past nine are both recorded as occurring at 09:05, then they appear in the record as simultaneous; but a more accurate clock would have been able to reveal that the first bid was made at 09:05:01.64 and the second at 09:05:01.98. Even if two events occur absolutely at the same time, very often some random process (e.g. what direction the auctioneer is looking in) acts to break the simultaneity.

Thus, we may simulate real marketplaces (and implement electronic marketplaces) using techniques where each significant event always occurs at a unique time. We may choose to represent these by real high-precision times, or we may abstract away from precise time-keeping by dividing time (possibly irregularly) into discrete *slices*, numbered sequentially, where one significant event is known to occur in each slice. In the ZIP-trader markets explored here, we use such a time-slicing approach. 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 Ras *et al.* (2001) used a continuous-time formulation of the ZIP-trader algorithm).

In the markets described here (and in Cliff, 1997; 1998; 2001a; 2001b), 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 $p_j(t)<=q(t)$ 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 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 a maximum number of time-slices have run.

B. Space of Possible Auctions

Now consider the case where we implement a ZIP-trader continuous double auction (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 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. Here, as in previous ZIP work (Cliff, 1997; 1998; 2001a; 2001b) the random binary variable is always independently and identically distributed over all time-slices.

So, 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 English Auction (EA) we have $Q_s=0.0$, and in the Dutch Flower Auction (DFA) we have $Q_s=1.0$. Thus, there are at least three values of Q_s (*0.0, 0.5,* and *1.0*) that correspond to three types of auction familiar from centuries of human economic affairs. 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.

The inventive step introduced by Cliff (2001b) was to consider the Q_s values of 0.0, 0.5, and 1.0 not as three distinct market mechanisms, but rather as the two end points and the midpoint on a *continuum* of mechanisms. For values other than these, there is a straightforward implementation. For example, Q_s=0.1 can be interpreted as specifying an auction mechanism where, 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 I am 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 $Q_s=0.712803$ (say) are preferred. Given the infinite nature of a real continuum, it seems appealing to use an automatic exploration process, such as the GA, to identify useful values of Q_s.

Thus, Cliff (2001b) added a ninth dimension to the search space, and the genotype in the GA became the eight real values for ZIP-trader initialisation, plus a real value for Q_s. As with all prior experiments (Cliff, 1998; 2001a; & 2001b), no "NYSE" improvement rule (Cliff, 1997) was used.

C. The Genetic Algorithm

The same simple GA used by Cliff (2001b) is used here, with one difference. Cliff (2001b) used a population of size 30 and evolution was allowed to progress for 1000 generations. Each experiment was repeated 50 times, and it was found that several of the experiments yielded multimodal results. However, in all the experiments reported on in that paper, the qualitative nature of the results was very clear by generation 500: all runs settled to a particular mode by generation 300, and the improvement in performance (i.e., fitness) between generation 500 and generation 1000 was always very small. Thus, all the experiments reported on in this paper ended after 500 generations. All other GA control parameters are unchanged. For an introduction to GAs, see Mitchell (1998).

In each generation, all individuals were evaluated and assigned a fitness value; and the next generation's population was then generated via mutation and crossover on parents identified using rank-based tournament selection. Elitism (where an unadulterated version of the fittest individual from each generation is copied into each successive generation) 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 ***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 number of generations (here $n_g=1000$ for consistency with Cliff (2001b), despite the fact that all experiments are now terminated after 500 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, as in (Cliff, 2001b), $m_s=0.05$ 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 has been shown by Bullock (1999) to bias 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 than values within those bounds. Moreover, initial and mutated genome values of μ_Δ, β_Δ, and γ_Δ were clipped where necessary to satisfy the constraints $(\mu_{min}+\mu_\Delta)<=1.0$, $(\beta_{min}+\beta_\Delta)<=1.0$, and $(\gamma_{min}+\gamma_\Delta)<=1.0$.

The fitness of genotypes was evaluated using the methods described by Cliff (1998; 2001a; 2001b). One *trial* of a particular genome was performed by initialising a ZIP-trader market from the genome, and then allowing the ZIP traders to operate within the market for a fixed number of trading periods, with allocations of stock and currency being replenished between each trading period. Each trading period ended either when no more trades was possible, or a maximum number of time-slices was reached.

During each trading period, Smith's (1962) α measure (root mean square deviation of transaction prices from the theoretical market equilibrium price) was monitored, and a weighted average of α was calculated across the trading periods in the trial as described in Section II.E below. 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: that is, we are attempting here to *minimise* the α fitness value.

D. Previous Results

In Cliff (2001b), results were presented from three evolution experiments. In each experiment, a single fixed market supply and demand schedule was used for every trial in the experiment. These three schedules are referred to as markets M1, M2, and M3. In all of them there were 11 buyers and 11 sellers, each empowered to buy/sell one unit of commodity: markets M1 and M2 are illustrated in Figures 1 and 2.

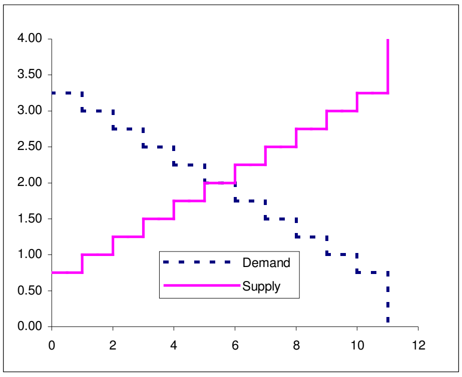

Figure 1: Supply and demand schedules for market M1

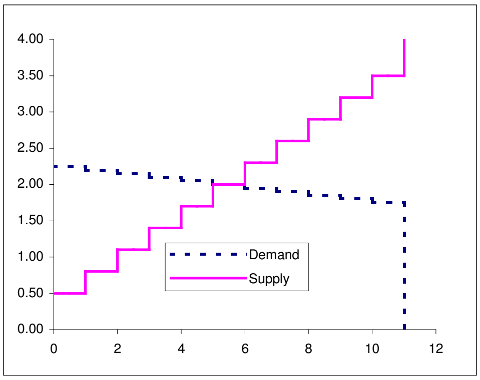

Figure 2: Supply and demand schedules for market M2.

The best mode of evolved results from experiments with M1 all had values of Q_s =~0.0. The best mode of evolved results from experiments with M2 all had values of Q_s that, while clearly nonzero, were sufficiently small that they did not give performance that was statistically better than if Q_s had been fixed at zero *a priori*. That is, with both M1 and M2 the evolved value of Q_s was effectively zero, and so the EA-like one-sided auction was found to be the most efficient market mechanism. For M3, however, values of Q_s around 0.16 consistently evolved, and this hybrid auction mechanism was shown to be more efficient than the previously-known Q_s=0.0 and Q_s=0.5 mechanisms that it lies between.

E. Dual-Schedule "Market-Shock" Experiments

Because for each trial in all three of Cliff's (2001b) experiments a single fixed market schedule was used in evaluating the evolving solutions, there is a manifest possibility that the GA tailored the final evolved solutions to peculiarities of the specific market schedules employed – i.e., that it "overfitted". To test this hypothesis, a new set of experiments were run, where "shock changes" are inflicted on the market by swapping from one schedule to another halfway through the evaluation process. For ease of comparison with the results presented by Cliff (2001b), a six-period duration was used for each schedule, meaning that one trial now lasts for 12 periods: six periods with the ZIP trading agents adapting to trade under the first schedule, then at the end of the sixth period a sudden shock change of the market supply and demand to the second schedule (without altering any of the traders' parameters or variable values) followed by six periods of the traders adapting to trade and under that new schedule.

In Cliff (2001b), the evaluation function was a weighted average of Smith's α measure: 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 $w_s=\Sigma w_p$. In the dual-schedule experiments reported here, this was simply extended so that p=1…12 and $w_{p>6}=w_{p-6}$.

Two sets of experiments are reported on here: one set is referred to as "M1M2" to denote the fact that market M1 was used for the first six periods followed by M2 for the second six periods; and the other is referred to as "M2M1" to denote the reverse situation.

In any one experiment, there are 30 individuals evaluated over 500 generations where each evaluation involves calculating the mean of one hundred 12-period trials, so a total of 1.5 million market trials would be executed in any one GA experiment (on a Hewlett-Packard Kayak XU800 workstation this would take approximately 5 hours). Nevertheless, the progress of each GA experiment is itself affected by stochasticity (e.g. the GA may become trapped on local minima) and so to generate reliable results each experiment was repeated 50 times (i.e., 75 million market trials, taking approximately 10.5 days). Results from eight such 50-repeat experiments are discussed here (i.e., 84 days processing on one machine).

III. RESULTS

A. M1M2

Figure 3 shows results from 50 repetitions of the M1M2 evolving-market (EM) experiment: for each experiment, the fitness of the best (elite) member of the population is recorded. The results are clearly bimodal. Of the 50 repetitions, in 36 the elite ends up on fitness minima of about 3.85, while the other two elite fitness mode involves less-good minima around 4.2 to 4.3. Figure 4 shows the evolutionary trajectory of the mean value of Q_s calculated over the 36 members of the best elite mode. Clearly, the elite mode uses a hybrid auction mechanism partway between the one-sided Q_s =0.0 market and the Q_s=0.5 CDA.

For comparison, Figures 5, 6, and 7 show the fitness values from 50 repetitions of the M1M2 experiment in fixed-market (FM) conditions (i.e., where the value of Q_s was *not* evolved) for Q_s=0.0, Q_s=0.5, and Q_s=1.0 respectively. Using Q_s=0.0 is plausible because in Cliff (2001b) separate experiments evolving on M1 and on M2 alone both converged on optima at Q_s=0.0. Moreover, using Q_s=0.5 gives a CDA, and the CDA is often applauded as an auction mechanism in which transaction-price equilibration is rapid and stable, so we could plausibly expect the best fitness from using that market type. Fixed-market Q_s=1.0 results in Figure 7 are included for completeness, as this is analogous to the human-designed DFA mechanism.

With Q_s fixed at zero, Figure 5 shows that the mean best-mode elite score is around 4.1; and with Q_s=1.0, the results are worse by a factor of more than two. With the fixed CDA Q_s=0.5 auction style, an average elite fitness of around 4.05 is settled on by almost all experiments, as shown in Figure 6. To ease the comparison between the EM and FM-CDA results, Figure 8 shows the mean and standard deviation of the best-mode elite scores on the same graph. The EM results are clearly lower (and hence better) than those for the FM CDA.

As our fitness values are effectively measures of market efficiency, from Figure 8 it appears that using Q_s values of 0.25 give more efficient markets than using the previously "known" Q_s values such as 0.0, 0.5, or 1.0 for the M1M2 schedule sequence.

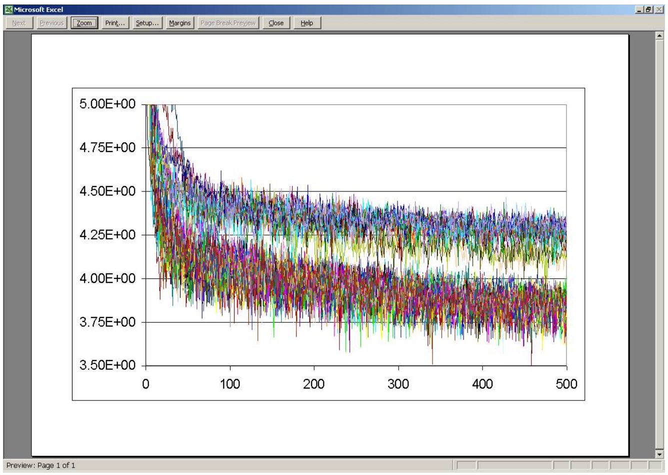

Figure 3: Elite fitness values from 50 repetitions of the 500-generation M1M2 evolving-market (EM) experiment. Lower values are better solutions (less deviation of transaction prices from the market's theoretical equilibrium price). Results are bimodal, with 36 of the repetitions (72%) settling in the best mode with values around 3.85.

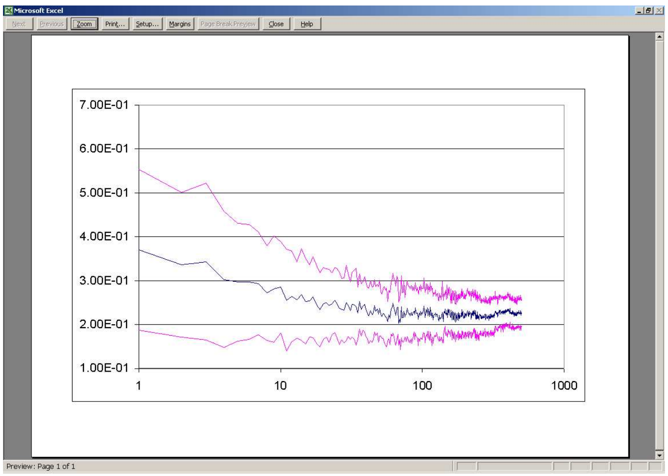

Figure 4: Evolutionary trajectory of mean (plus and minus one s.d.; n=36) value of Q_s in the best elite mode of the 50 M1M2 EM experiments shown in Figure 3. The mean settles to approximately 0.25

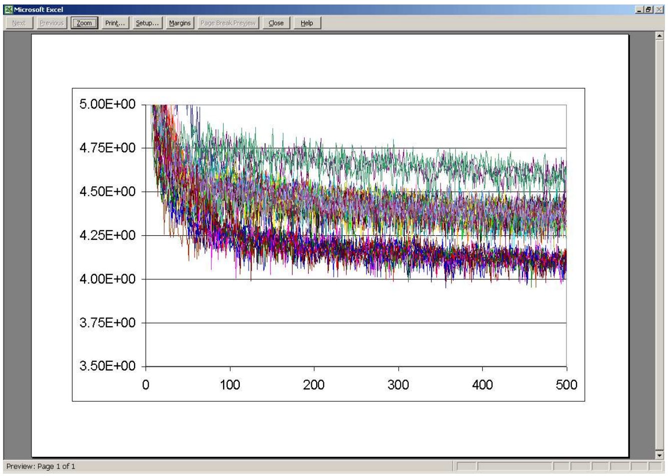

Figure 5: Elite fitness values from 50 repetitions of a 500-generation M1M2 fixed-market (FM) experiment with Q_s=*0.0*. Results are trimodal, with the 9 experiments (18%) in the best mode settling to values around 4.1.

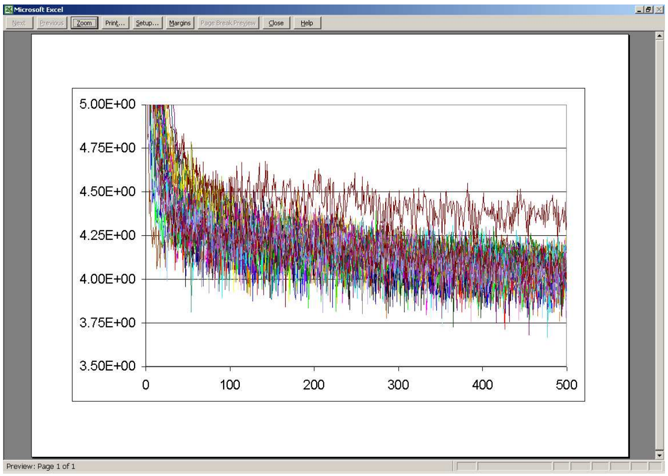

Figure 6: Elite fitness values from 50 repetitions of a 500-generation M1M2 fixed-market (FM) experiment with Q_s=*0.5*. Results are bimodal, with 49 of the repetitions (98%) settling in the best mode with values around 4.05

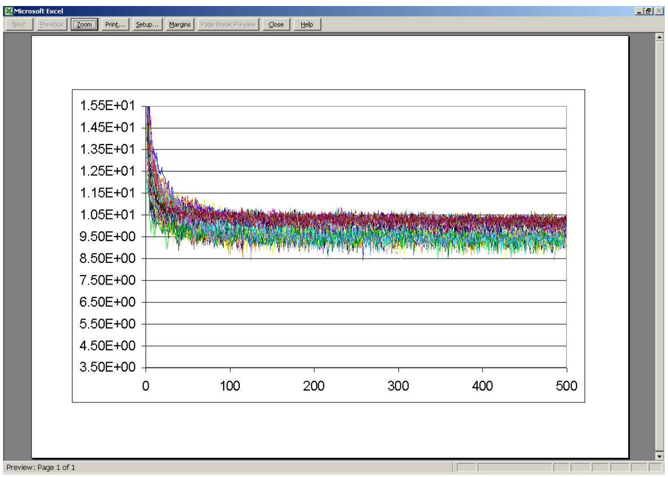

Figure 7: Elite fitness values from 50 repetitions of a 500-generation M1M2 fixed-market (FM) experiment with Q_s=*1.0*. Results are bimodal, with both modes in the range 8.5 to 10.5.

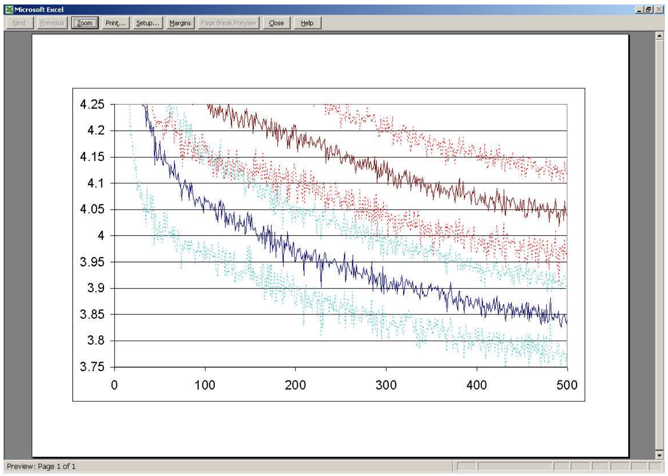

Figure 8: Average elite fitnesses from 50 EM and 50 FM(Q_s=*0.5*) M1M2 experiments; data is plotted for mean fitness, plus and minus one standard deviation (s.d.). Best EM fitness mode settles to a mean of approx 3.85 with a s.d. of approx 0.06 (*n*=36); FM values settle to a mean of around 4.05 with a s.d. of approx. 0.1 (*n*=49).

This Q_s=0.25 mechanism could easily be implemented in an electronic marketplace by allowing, on the average, one quote in four to come from a seller while the remaining quotes come from buyers.

B. M2M1

Figure 9 shows the evolutionary trajectories for the elite fitness scores in 50 repetitions of the M2M1 EM experiment. The best mode has a mean fitness of 4.18. Figure 10 shows the evolutionary trajectory of the mean Q_s value in the best mode: a final value of approximately 0.45 is settled on. This is sufficiently close to the CDA value of Q_s=0.5 to arouse suspicion that the EM results have settled to a Q_s value that yields results statistically indistinguishable from those that would result if a CDA mechanism had been chosen *a priori*. To test this hypothesis, 50 repetitions of a FM experiment with Q_s=0.5 were run, and the elite fitness results are illustrated in Figure 11.

The FM-CDA (Q_s=0.5) results for M2M1 shown in Figure 11 are unimodal, and the mean elite fitness is again approx 4.18. Thus, simple visual comparison (*cf.* Figure 8) is not sufficient to establish any statistically significant difference between best modes shown in the M2M1 EM and FM-CDA results.

Following Cliff (2001b), the Wilcoxon version of the Wilcoxon-Mann-Whitney Test (Siegel & Castellan, 1988) was used to see if there is a statistically significant difference between the M2M1 EM and FM(Q_s=*0.5*) results. The *m*=50 final best-mode fitness scores from the FM(Q_s=*0.5*) experiment and the *n*=46 final best-mode fitness scores from the EM experiment were grouped together, with the EM values marked as Type 1 and the FM values marked as Type 2. Fitness values were then assigned a rank-order based on their position following sorting into ascending order. There were no tied ranks. Summing the rank values for Type 2 (FM) gave a value W_2=2480. Using this value and $N=m+n$ for $z=(W_2+0.5-m(N+1)/2)/(mn(N+1)/12)^{0.5}$ gives z=0.407037, which is not significant. Thus it can be concluded that there is no statistically significant difference between the results from the EM and FM(Q_s=0.5) experiments for M2M1.

Furthermore, it should be noted that, as with the M1M2 schedule ordering, the FM(Q_s=0.0) and FM(Q_s=1.0) results for M2M1 were clearly and consistently worse than either the EM or FM(Q_s=0.5) results: see Figures 12 and 13 respectively.

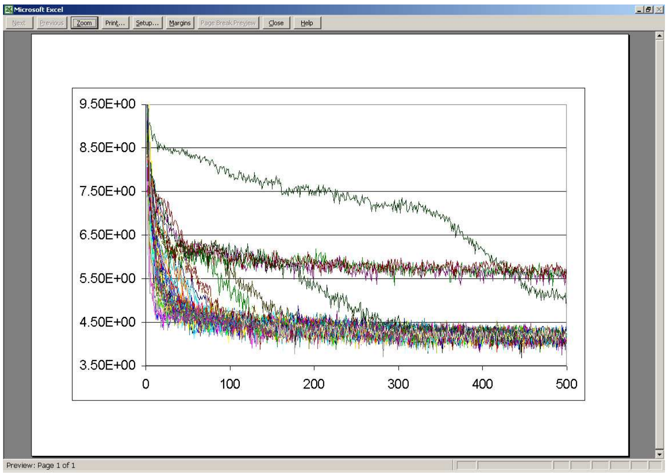

Figure 9: Elite fitness values from 50 repetitions of the 500-generation M2M1 EM experiment. There are two clear modes: one (n=3) at approx 5.6 and the other (n=46) around 4.18. The one "outlier" result may be heading for the lower mode or may be a distinct third mode.

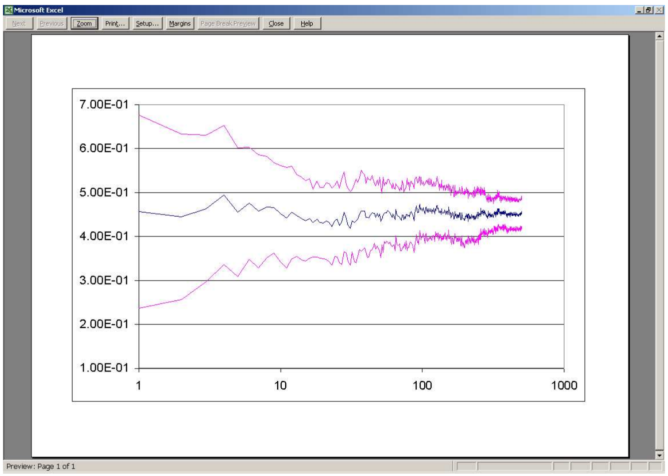

Figure 10: Evolutionary trajectory of mean (plus and minus one s.d.; n=46) value of Q_s in the best elite mode of the 50 M2M1 EM experiments shown in Figure 9. The mean settles to approximately 0.45

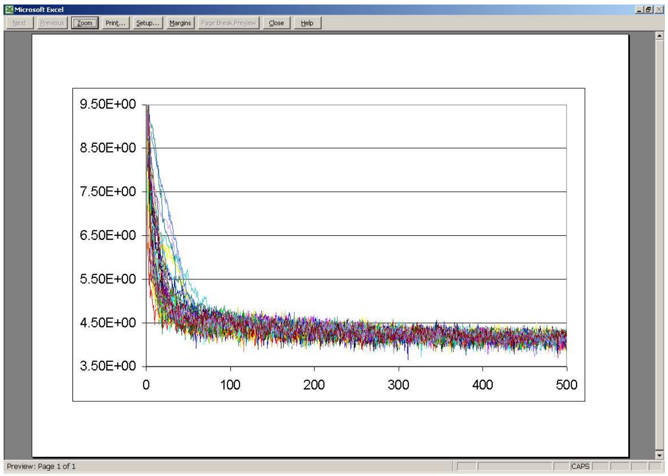

Figure 11: Elite fitness values from 50 repetitions of a 500-generation M2M1 fixed-market (FM) experiment with Q_s=0.5. Results are unimodal, with all the repetitions settling in the best mode with values around 4.2

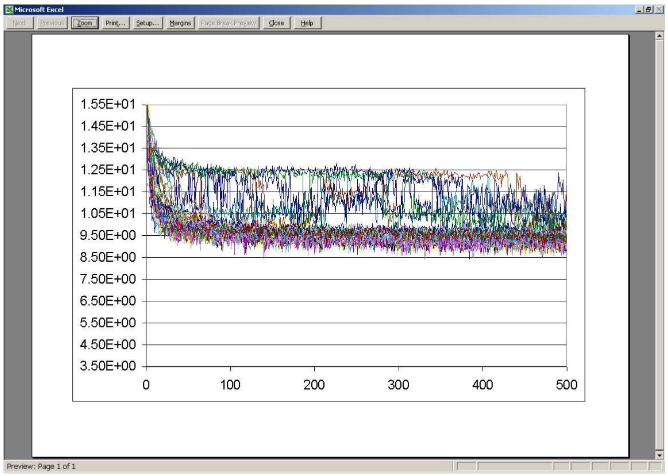

Figure 12: Elite fitness values from 50 repetitions of a 500-generation M2M1 fixed-market (FM) experiment with Q_s=0.0. For much of the experiment the results are multi-modal, but by generation 500 all the repetitions settle with values around 9 to 11.

IV. DISCUSSION AND CONCLUSION

Note that although one-sided Q_s=~0.0 mechanisms were evolved in (Cliff, 2001b) for M1and M2 individually, when the traders have to deal with the "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 the M1M2 experiments 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 of Q_s=0.5. Thus, this paper extends the line of research first reported on by Cliff (2001b). It again demonstrates 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. And again new "hybrid" market mechanisms were found to give better market dynamics for M1M2 than the previously-known auction styles. Once again, while such evolved market mechanisms are unlike any human-designed mechanism, they could nevertheless readily be implemented as online electronic marketplaces.

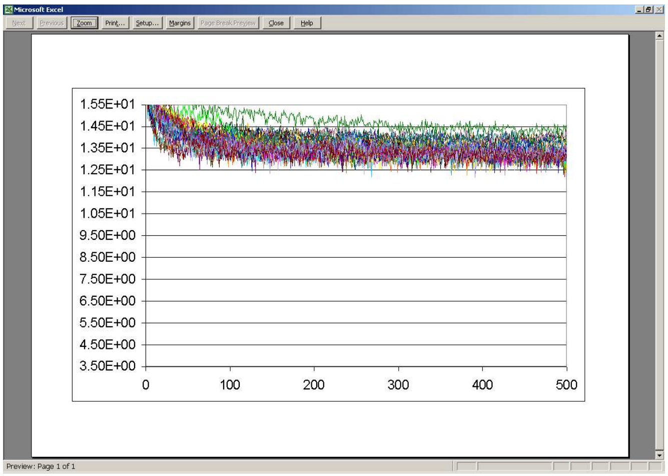

Figure 13: Elite fitness values from 50 repetitions of a 500-generation M2M1 fixed-market (FM) experiment with Q_s=*1.0*. Results are unimodal, with all the repetitions settling to values around 13 to 14.

Thus, one contribution of this paper is the demonstration that the Q_s=*~0.0* results for M1 and M2 in Cliff (2001b) are consequences of (unrealistically) using unchanging supply and demand curves for the duration of each experiment. The results presented here show that, for dealing with shock changes from M1 to M2, and from M2 to M1, Q_s=*~0.0* is not the best value, even though it was the optimum for each market individually. A second contribution is the demonstration that the optimum Q_s value is order-dependent: i.e., that the evolved value of Q_s for M1M2 is different to that for M2M1.

It is widely acknowledged within artificial evolution research that blind evolutionary search processes such as that implemented by the GA used here will frequently improve fitness via ruthless exploitation of any regularity in the task environment. We have seen that, although in the M2M1 experiments no such regularity was identified for exploitation, in the M1M2 experiments there was an underlying regularity that allowed an evolved hybrid (Q_s=*0.25*) market mechanism to be more efficient. Thus, the major contribution of this paper is to demonstrate that, even when there are shock changes in supply and demand, there may be sufficient regularity in some market situations such that non-CDA hybrid two-sided auctions are more efficient than any human-designed market mechanism. Given these results, coupled with the results of Das *et al.* (2001) who demonstrated that ZIP artificial 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 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 in major international financial markets (populated by artificial trading agents) will be savings or profits measured in billions of dollars.

Acknowledgements

Thanks once again to Rycharde Hawkes for his co-operation with my use of an ad-hoc compute farm. Thanks also to the HP Labs Biologically-Inspired Complex Adaptive Systems (BICAS) group for valuable discussions. See www.hpl.hp.com/research/bicas.

References

S. Bullock (1999), "Are artificial mutation biases unnatural?" in: Floreano, D., Nicoud, J.-D. & Mondada, F. (eds) *Advances in Artificial Life - Fifth European Conference (ECAL99)*, pp. 64-73 Springer-Verlag. 1999. http://www.comp.leeds.ac.uk/seth/papers/ecal99_mutation.ps.gz

D. Cliff (1997), "Minimal-intelligence agents for bargaining behaviours in market environments" Technical Report HPL-97-91, Hewlett-Packard Labs. http://www.hpl.hp.com/techreports/97/HPL-97-91.html

D. Cliff (1998), "Genetic optimization of adaptive trading agents for double-auction markets" in *Proceedings of Computational Intelligence in Financial Engineering (CIFEr) 1998.* IEEE/IAFE/Informs (preprint proceedings), pp.252-258, 1998.

D. Cliff (2001a), "Evolutionary optimization of parameter sets for adaptive software-agent traders in continuous double-auction markets", Technical Report HPL-2001-99, Hewlett-Packard Laboratories. http://www.hpl.hp.com/techreports/2001/HPL-2001-99.html

D. Cliff (2001b), "Evolution of market mechanism through a continuous space of auction-types". To be presented at *Computational Intelligence in Financial Engineering* session at *CEC2002*, Hawaii, May 2002. Technical Report HPL-2001-326: Hewlett-Packard Laboratories. http://www.hpl.hp.com/techreports/2001/HPL-2001-326.html

R. Das, J. Hanson, J. Kephart, & G. Tesauro (2001),"Agent-human interactions in the continuous double auction" *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-01)*, 2001. http://www.research.ibm.com/infoecon/researchpapers.html

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

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

M. Mitchell (1998). *An Introduction to Genetic Algorithms.* MIT Press.

NYSE (2002), *Stock Market Activity* report available at: http://www.nyse.com/pdfs/02_STOCKMARKETACTIVITY.pdf

S. Siegel & N. J. Castellan (1988), *Nonparametric Statistics for the Behavioural Sciences*. McGraw Hill.

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