

## **Explorations in Evolutionary Design of**

### **Online Auction Market Mechanisms**

*Dave Cliff*

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

dave.cliff@hp.com    phone +44 117 312 8189    fax +44 117 312 8924

#### **Abstract**

This paper describes the use of a genetic algorithm (GA) to find optimal parameter-values for trading agents that operate in virtual online auction “e-marketplaces”, where the rules of those marketplaces are *also* under simultaneous control of the GA. The aim is to use the GA to automatically design new mechanisms for agent-based e-marketplaces that are more efficient than online markets designed by (or populated by) humans. The space of possible auction-types explored by the GA includes the Continuous Double Auction (CDA) mechanism (as used in most of the world’s financial exchanges), and also two purely one-sided mechanisms. Surprisingly, the GA did not always settle on the CDA as an optimum. Instead, novel *hybrid* auction mechanisms were evolved, which are unlike any existing market mechanisms. In this paper we show that, when the market supply and demand schedules undergo sudden “shock” changes partway through the evaluation process, two-sided hybrid market mechanisms can evolve which may be unlike any human-designed auction and yet may also be significantly more efficient than any human designed market mechanism.

**Keywords:** Online Auction Marketplaces; e-Marketplaces; Automated Market Mechanism Design; Trader-Agents; ZIP Traders; Genetic Algorithms.

## 1. Introduction

For thousands of years, buyers and sellers have come together to exchange money for goods or services. Economists use the word “auction” to refer to the mechanism (or rules) by which buyers and sellers interact in such marketplaces. Almost all traders in the global international financial markets interact via a particular form of auction market mechanism known as the *continuous double auction* (CDA), more details of which will be given later.<sup>1</sup> The CDA has been the subject of much study by economists, partially because it is so important in the world of finance, but also because CDA markets typically exhibit a very attractive characteristic: experimental studies have demonstrated that the transaction prices in a CDA market rapidly converge on the market’s theoretical *equilibrium price*. Students of microeconomics know the equilibrium price as the price at which the market’s supply and demand curves intersect; but, colloquially, the equilibrium price is important because if transactions are taking place at off-equilibrium prices then someone somewhere in the market is being ripped off. Hence, rapid equilibration is desirable in any auction. The precise reasons why CDA markets typically exhibit rapid and stable equilibration are still the topic of research and debate (see e.g. [12]).

With the advent of e-commerce, various forms of auction mechanism have become very popular for online trading, and web-based auction sites such as [www.ebay.com](http://www.ebay.com) have proven highly successful. As auctions dematerialize, moving online and becoming virtual “e-marketplaces”, it becomes perfectly plausible for software-agent “robot” traders to participate in those auctions. In comparison to human traders, such “bots” have the

---

<sup>1</sup> It is beyond the scope of this paper to provide a review of all possible auction mechanisms: for a recent relevant paper, see [25].

advantage of being very fast and very cheap, and in principle they can assimilate and act on volumes of data that would swamp even the most able of human traders.

ZIP (Zero-Intelligence-Plus) artificial trading agents, introduced in [3], are software-agent “trader bots” 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 Smith’s [22] pioneering experimental economics studies of the CDA and other auction mechanisms. ZIP traders were originally developed as a solution to the pathological failures of Gode & Sunder’s (1993) “ZI” (Zero-Intelligence) traders, but recent work at IBM by Das *et al.* [11] has shown that ZIP traders (unlike ZI traders) consistently outperform human traders in human-against-robot experimental economics CDA marketplaces. The ZIP traders consistently made profits a few percentage points higher than did the human traders they were competing against. Das *et al.* [11] wrote that the “...successful demonstration of machine superiority in the CDA ... could have a ... powerful financial impact – one that might be measured in billions of dollars annually”, and in their conclusions they speculate on the future possibility of online e-marketplaces currently populated by human traders becoming populated entirely by trader agents.

The operation of ZIP traders has been successfully demonstrated in experimental versions of CDA markets similar to those found in the international financial markets for commodities, equities, capital, and derivatives; and in posted-offer auction markets similar to those seen in domestic high-street retail outlets [3]. In any such market, there are a number of numeric parameters that govern the adaptation and trading processes of the ZIP traders. In the original 1997 version of ZIP traders, the values of these were set

by hand, using “educated guesses”. However, subsequent papers [4,5] presented the first results from using a standard technique to automatically optimize these parameter values, thereby eliminating the need for skilled human input in deciding the values.

Prior to the research described in [6], 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, in the absence of an auctioneer).

In this paper, we explore in detail the some specific consequences of asking the following question: if, as Das *et al.* [11] speculate, trader agents will come to replace human traders in online e-marketplaces, then why should those online e-marketplaces use auction mechanisms designed by humans, for humans? Perhaps there are new market mechanisms, *suitable only to populations of robot-traders*, that are more efficient (or otherwise more attractive) than currently-known human-based mechanisms.

Designing new market mechanisms is hard, and the space of possible mechanisms is vast. For this reason it is attractive to use an *automated search* of the space of possible mechanisms: in essence, we ask a computer to do the auction-design for us. This paper reports on exploring the application of one type of automated search/optimization algorithm, which is inspired by Darwinian notions of evolution via random variation and directed selection, and hence is known as a Genetic Algorithms (GA).

The first results from experiments where a GA optimizes not only the parameter values for the ZIP trading agents, but also the style of market mechanism in which those traders operate, were presented in [6]. 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 an automated search through a continuous space of possible auction-types. Thus, the results in [6] 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 (or even only a few basis 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 [16]). 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 be made at NASDAQ, at European exchanges such as LSE and LIFFE, and at similar exchanges elsewhere around the globe.

Section 2 gives an overview of ZIP traders and of the experimental methods used, including a description of the continuously-variable space of auction types. This description is largely identical to the account given in previous papers [6,7], albeit extended to describe how the new experiments whose results are presented here differ from the previous work. The new results are presented in Section 3 and are discussed in Section 4. Related work is reviewed in Section 5, and conclusions are drawn in Section 6.

## **2. Methods**

### **2.1 Zero-Intelligence-Plus (ZIP) Traders**

ZIP trading agents are described fully in a lengthy report [3], 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.

ZIP traders deal in arbitrary abstract commodities. Each ZIP trader  $i$  is given a private (i.e., secret) limit-price,  $\lambda_i$ , 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  $\lambda_i$  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 utility 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 maximize their utility over all trades, where utility is the difference between the accepted quote-price and the trader’s  $\lambda_i$  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 [20] and in learning classifier systems [24]. This rule has a “learning rate” parameter  $\beta_i$  that governs the speed of convergence between trader  $i$ ’s quoted price  $p_i(t)$  and the trader’s idealized “target” price  $\tau_i(t)$ . When calculating  $\tau_i(t)$ , traders introduce a small random absolute perturbation generated from<sup>2</sup>  $U[0, c_a]$  (this perturbation is positive when increasing  $\tau_i(t)$ , negative when decreasing) and also a small random relative perturbation generated from  $U[1-c_r, 1]$  (when decreasing  $\tau_i(t)$ ) or  $U[1, 1+c_r]$  (when increasing  $\tau_i(t)$ ). Here  $c_a$  and  $c_r$  are global system constants. To smooth over noise in the learning system, there is an additional “momentum” parameter  $\gamma_i$  for each trader (such momentum terms are also commonly used in back-propagation neural networks).

---

<sup>2</sup> Note that in this paper  $v=U[x,y]$  denotes a random real value  $v$  generated from a uniform distribution over the range  $[x,y]$ .

Thus, adaptation in each ZIP trader  $i$  has the following parameters: initial margin  $\mu_i(0)$ ; learning rate  $\beta_i$ ; and momentum term  $\gamma_i$ . In an entire market populated by ZIP traders, values for these three parameters are randomly assigned to each trader via the following expressions:  $\mu_i(0)=U(\mu_{min}, \mu_{min}+\mu_{\Delta})$ ;  $\beta_i=U(\beta_{min}, \beta_{min}+\beta_{\Delta})$ ; and  $\gamma_i=U(\gamma_{min}, \gamma_{min}+\gamma_{\Delta})$ .

Hence, to initialize an entire ZIP-trader market it is necessary to specify values for the six market-initialization parameters  $\mu_{min}$ ,  $\mu_{\Delta}$ ,  $\beta_{min}$ ,  $\beta_{\Delta}$ ,  $\gamma_{min}$ , and  $\gamma_{\Delta}$ ; and also for the two global system constants  $c_a$  and  $c_r$ . And so it can be seen that any set of initialization parameters for a ZIP-trader market exists within an eight-dimensional real space. Vectors in this 8-space can be considered as “genotypes” in a genetic algorithm (GA), 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 in [4,5], and that is described further below. Before that, the issue of simulating the passage of time is discussed.

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 to be simultaneous; but a more accurate clock would have been able to reveal that the first bid was made at 09:05:01 and the second at 09:05:02. Even if two events occur absolutely at the same time, some random process (e.g. what direction the auctioneer is looking in) may 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 Das *et al.* [11] used a continuous-time formulation of the ZIP-trader algorithm).

In the markets described here (and in [3,4,5,6,7,8,9]), 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  $p_j(t) \leq q(t)$  for sellers, or if  $p_j(t) \geq q(t)$  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  $\lambda_i$  limit prices, until either a maximum number of time-slices have run, or a maximum number of sequential quotes have been ignored.

## 2.2. A 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 next 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 [3,4,5,6,7,8,9] 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 in [6] 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 endpoints 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 of such markets: why would anyone suggest such a bizarre way of operating? And who would go to the trouble of setting themselves up to act as an auctioneer for such a mechanism? Certainly, it is perfectly possible for a human auctioneer to run an auction using a value of  $Q_s$  other than 0.0, 0.5, or 1.0. For any given value of  $Q_s$ , all that the auctioneer needs is an unbiased roulette-wheel partitioned into two segments: one marked “Seller” and measuring  $Q_s*360$  degrees of arc; the other marked “Buyer” and measuring  $(1.0-Q_s)*360$  degrees of arc. To determine the source of each successive new quote in the auction, the auctioneer would spin the wheel and then, depending on whether the ball ends up in the “Seller” or the “Buyer” segment, would take the next quote either from a seller or a buyer. Clearly, an online version of such an auction mechanism can be implemented in only a few lines of code, so long as an appropriate method for generating random numbers is available. But (to the best of my knowledge) neither the manual roulette-wheel version nor the online implementation of such auction mechanisms have ever been implemented before for any value of  $Q_s$  other than 0.0, 0.5, or 1.0.

Nevertheless, there is no *a priori* reason to argue that these three previously-known points on this  $Q_s$  continuum are the only loci of useful auction types. Maybe there are circumstances in which values such as  $Q_s=0.25$  (say) are preferred. Given the infinite nature of this real continuum it seems appealing to use an automatic exploration process, such as the GA, to identify useful values of  $Q_s$ .

Thus, in [6] a ninth dimension was added to the search space, and the genotype in the GA became the eight real values for ZIP-trader initialization, plus a real value for  $Q_s$ . No “NYSE” quote-improvement rule [3] was used in the experiments reported in this paper.

### **2.3. The Genetic Algorithm**

The simple GA used in [5] is also used here, with one difference. In [5] a population of size 30, evolving for 1000 generations, was used. Each experiment was repeated 50 times, and it was found that several of the experiments yielded multi-modal results. However, in all the experiments reported on in that paper, the qualitative nature of the outcome of the experiment 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 the experiments reported here were ended after 500 generations. All other GA control parameters are unchanged. For an introduction to GAs, see [14] or [15].

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, on each generation, an unadulterated version of the fittest individual from the evaluated population is copied into the new successor population) 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  $U[-m(g), +m(g)]$  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:  $\log_{10}(m(g)) = \log_{10}(m_s) - ((g/(n_g-1)) \cdot \log_{10}(m_s/m_e))$  where  $n_g$  is the number of generations (here  $n_g=1000$  for consistency with [6], 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 [6],  $m_s=0.05$  and  $m_e=0.0005$ .

If ever mutation caused the value at a locus to fall outside the range  $[0.0, 1.0]$  it was simply clipped to stay within that range. This clip-to-fit approach to dealing with out-of-range mutations has been shown [1] 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  $\mu_\Delta$ ,  $\beta_\Delta$ , and  $\gamma_\Delta$  were clipped where necessary to satisfy the constraints  $(\mu_{min} + \mu_\Delta) \leq 1.0$ ,  $(\beta_{min} + \beta_\Delta) \leq 1.0$ , and  $(\gamma_{min} + \gamma_\Delta) \leq 1.0$ .

The fitness of genotypes was evaluated using the methods described in [4,5,6]. One *trial* of a particular genome was performed by initializing 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. During each trading period, Smith's [22]  $\alpha$  measure (root mean square deviation of transaction prices from the theoretical market equilibrium price) was monitored, and a weighted average of  $\alpha$  was calculated across the trading periods in the trial, using the method described in Section 2.5 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: we aim to *minimize* fitness scores.

#### **2.4. Previous Results**

Results from nine investigative sets of experiments have been presented in our prior publications. Those results are included for completeness in the tables presented in Section 3, where results from an additional 23 new experiments are published for the first time. All the experiments whose results are tabulated in Section 3 involve evaluating the performance of the evolving auction-market mechanisms on one or more of four market supply and demand schedules. These four schedules are referred to as markets M1, M2, M3, and M4, and are illustrated in Figure 1.

\*\*\* FIGURE 1 NEAR HERE \*\*\*

In all four schedules there are 11 buyers and 11 sellers, each empowered to buy/sell one unit of commodity: these relatively small numbers are the cause of the stepped supply and demand curves. Market M1 is taken from [22]. The remaining three markets are minor variations on M1. In M2 the slope of the demand curve has been greatly reduced

while the slope of the supply curve has been increased only slightly; and in M4 the slope of the supply curve has been greatly reduced while the slope of the demand curve has been increased only slightly. In M3 the slopes of both the supply and demand curves are only slightly steeper than the slopes in M1. Despite the apparent similarity between M1 and M3, a detailed empirical study presented in [8] demonstrated that these minor differences between the supply and demand curves in M1 and M3 can lead to significant differences in the final best evolved solutions.

In the so-called “single-schedule” experiments, only one of the market schedules was used throughout the evolutionary process. Results from the four single-schedule experiments are summarized in Table 1 in Section 3. The key qualitative issue is that in all four experiments, the best evolved mechanisms all differed from the CDA, and in two cases the best evolved mechanism was not even a one-sided auction like the EA or DFA mechanisms; rather, the best evolved auction-mechanism was a peculiar hybrid, partway between the CDA and a pure one-sided auction.

However, because for each trial in all four of these 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 supply and demand schedules employed – i.e., that it “over-fitted”. To test this hypothesis, a new suite of experiments was run, where “shock changes” were inflicted on the market by swapping from one schedule to another partway through the evaluation process. The results from 19 of these experiments are presented in Section 3. Initially, dual-schedule experiments were run, where the supply and demand schedules were suddenly changed

halfway through the evaluation process. Some early results from these experiments were presented in [7]: these showed that when M1 was used for the first half of the evaluation, followed by M2 for the second half (which we refer to here as M1-2), the results evolved by the GA were order-dependent. That is, when the order of the schedules was reversed, so that in the evaluation process M2 was followed by M1 (which we refer to here as M2-1), the results differed from the M1-2 case. Furthermore, for both M1-2 and M2-1, the optimal evolved values of  $Q_s$  differed from the values that were found to be optimal when evaluation involved either M1 or M2 alone. The M1-2 results are presented in detail in the next section, as illustration of the process used to compare the results from evolving-mechanism (EM) experiments with the results from fixed-mechanism (FM) experiments. In all the FM experiments, the value of  $Q_s$  is not evolved, but the remaining eight ZIP-trader parameter-values on the genotype are still optimised by the GA. The M2-1 results are presented in summary form in Table 2 (Section 3), along with results from new dual-schedule experiments, presented for the first time in this paper.

The order-dependence shown by the M1-2 and M2-1 results could again potentially be a consequence of the GA over-fitting: a “dual schedule” experiment could also reasonably be described as a “single-shock” experiment; and perhaps the GA evolved solutions that were over-fitted to each particular shock. For instance, in the M1-2 case the GA might be over-fitting the evolved parameter-values and market-mechanism to the *specific* market-shock of suddenly transitioning from M1 to M2. To explore this possibility, additional sets of experiments were run where two shocks occurred during each evaluation process (i.e., switching between three schedules). Results from four such sets of triple-schedule experiments were presented in [9], all involving schedules M1, M2, and M3. In one



experiment, referred to here as M1-2-1, the evaluation involved six trading periods with supply and demand determined by M1, then a sudden change to M2, then six periods later a reversion to M1 for a final six periods. The other sets of experiments are referred to here as M2-1-2, M1-2-3, and M3-2-1 (and so on), the meaning of which should be obvious. The results from these four sets of experiments are presented in summary form in Table 3 (Section 3) along with results from an additional 14 new sets of experiments. For ease of comparison with the single-schedule results presented in [6], a six-period duration was used for each market schedule, meaning that a dual-schedule trial lasts for 12 periods: 6 periods with the ZIP trading agents adapting to trade under the first schedule, then at the end of the 6th 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 6 periods of the traders adapting to trade and under that new schedule. Similarly, the triple-schedule experiments each lasted for 18 trading periods.

In [6], the evaluation function was a weighted average of Smith’s  $\alpha$  measure: in each trading period  $p$  the value  $\alpha_p$  was calculated, and the fitness score was computed as  $(1/\sum w_p) \cdot \sum (\alpha_p \cdot w_p)$  for  $p=1 \dots 6$  with weights  $w_1=1.75$ ,  $w_2=1.5$ ,  $w_3=1.25$ , and  $w_{p>3}=1.0$ . In the dual-schedule experiments reported here, this was extended so that  $p=1 \dots 12$  and  $w_{p>6}=w_{p-6}$ . Similarly, in the triple-schedule experiments,  $p=1 \dots 18$  and  $w_{p>12}=w_{p-12}$ .

### 3. Results

Results from 32 sets of experiments are presented here: one set for each sequence of schedules explored. Each set involves 100 individual experiments: 50 repetitions of the GA experiment for the evolving-mechanism (EM) case where the value of  $Q_s$  is under

evolutionary control, and (for comparison) a further 50 repetitions for the same sequence in the fixed-mechanism (FM) case, where  $Q_s$  is fixed at the CDA value of 0.5. Of the 32 sets, 4 are single-schedule, 10 are dual-schedule and the remaining 18 are triple-schedule.

Section 3.1 gives a detailed presentation of results from the M1-2 case, for illustration of the process used to compare the EM and FM cases. Section 3.2 then presents tables summarizing the results from all the experiments performed so far.

### **3.1 Detailed Dual-Schedule Results: M1-2**

Figure 2a shows the fitness of all 30 genotypes in the population at each generation from 1 to 500 in a single run of the M1-2 evolving-market (EM) experiment. In each generation the elite (best-scoring) individual is of most interest, and Figure 2b shows the trajectory of the elite fitness score for the population shown in Figure 2a. The results shown in Figure 2 are non-deterministic: different runs of the GA (with different seed values for its random number generator) will yield different elite trajectories.

\*\*\* FIGURE 2 NEAR HERE \*\*\*\*

Examining the results from 50 repetitions of this experiment (with a different random seed used in each experiment), 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 3 shows the evolutionary trajectory of the mean and standard deviation (s.d.) of the  $Q_s$  values on the genomes of 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.

\*\*\* FIGURE 3 NEAR HERE \*\*\*\*

For comparison, similar trajectories of fitness values were recorded from 50 repetitions of the M1-2 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 [6] 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, which is often celebrated 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 were generated as this is analogous to the human-designed DFA mechanism.

With  $Q_s$  fixed at zero, 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 [7]. With the fixed CDA  $Q_s=0.5$  mechanism, an average elite fitness of around 4.05 is settled on by almost all experiments. To ease the comparison between the EM and FM-CDA results, Figure 4 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.

\*\*\* FIGURE 4 NEAR HERE \*\*\*\*

As our fitness values are effectively measures of market efficiency, from Figure 4 it appears that using  $Q_s$  values of around 0.23 give more efficient markets than using the previously “known”  $Q_s$  values such as 0.0, 0.5, or 1.0 for the M1-2 schedule sequence.

Noting that the evolved value of  $Q_s \approx 0.23$  is close to  $1/4$ , we can informally claim that a close approximation of this evolved auction mechanism could easily be implemented in an electronic marketplace by allowing, on the average, roughly one quote in four to come from a seller while the remaining three quotes in four come from buyers.

### 3.2 Summary Statistics

Having discussed the M1-2 results in detail, the tables in this section show summary data for a further 31 sets of experiments (each set consisting of 50 EM experiments and 50 FM experiments). As was stated earlier, results for M1, M2, and M3 were presented in [6]; Table 1 summarizes those results and presents new results from M4. The column labeling for all tables in this paper is as follows. The left-most column indicates the market schedules for each row of data. The column labelled “EM: $\mu$ ” is the mean fitness at generation 500 in the best elite mode from the 50 repetitions of the EM (evolving-market) experiment, and the column labelled “EM: $\sigma$ ” is the standard deviation for that mean. The column labelled “EM:n” shows the number of repetitions of the EM experiment that settled on the best elite fitness mode. The columns marked “FM: $\mu$ ”, “FM: $\sigma$ ”, and “FM:n” show the mean fitness, standard deviation of the mean fitness, and number of repetitions (from a total of 50) for the best elite fitness mode at generation 500 in the FM (fixed-market) experiments for each schedule. The column labelled “1%?” shows whether the Wilcoxon-Mann-Whitney test [21] indicates a statistically significant difference at the 1% confidence level between the EM and FM data. Finally, the columns labelled “ $Q_s$ : $\mu$ ” and “ $Q_s$ : $\sigma$ ” respectively show the mean  $Q_s$  value at generation 500, and the standard deviation on that mean, for the best elite mode from the EM experiments. Rows typeset

in italics are those for which there is a statistically significant difference at the 1% level between the EM and FM best elite mode data.

Results for M1-2 and M2-1 were previously presented in [7]; Table 2 summarizes those results and presents new results from an additional 8 single-shock experiments. Results for M1-2-1, M2-1-2, M3-2-1 and M1-2-3 were first presented in [9]; results from an additional 14 sets of dual-shock experiments are presented for the first time here in Tables 3 to 6.

\*\*\* TABLE 1 NEAR HERE \*\*\*

\*\*\* TABLE 2 NEAR HERE \*\*\*

Tables 3 to 6 all involve dual-shock (triple-schedule) evaluations, but they are grouped by the nature of the shocks. Table 3 shows results from experiments where only the demand curve undergoes a major change on each shock. Table 5 shows results from experiments where only the supply curve undergoes a major change on each shock. In Table 4, one of the two shocks involves a major change only to the demand curve while the other shock involves a major change only to the supply curve; and in Table 6 each shock involves a major change to both the supply curve and the demand curve.

\*\*\* TABLE 3 NEAR HERE \*\*\*

\*\*\* TABLE 4 NEAR HERE \*\*\*

\*\*\* TABLE 5 NEAR HERE \*\*\*

\*\*\* TABLE 6 NEAR HERE \*\*\*

Comparing the data in Tables 1 to 6, three points stand out. First, it is noticeable that in some cases, the elite evolved value of  $Q_s$  may differ quite markedly from the CDA value of 0.5, without there being a statistically significant effect on the market dynamics (i.e. on the fitness scores) in comparison to the FM  $Q_s=0.5$  case. For example, in both M2-1 and in M2-3-2 the EM  $Q_s$  values have a mean that is over two standard deviations away from the CDA value of 0.5, which on face value could lead one to expect that the mean EM and FM fitness scores would be significantly different; yet they are not. This is a consequence of the optimum  $Q_s$  value lying on a shallow plateau-like surface in the fitness landscape, such that apparently quite different values of  $Q_s$  yield very similar fitness values: a point explored and illustrated in detail in [8].

The second notable point is that the no-shock and single-shock data are not obviously useful in predicting the results of the dual-shock experiments, despite the fact that each of the dual-shock sequences explored in Tables 3 to 6 can be considered as the concatenation of two of the single-shock sequences explored in Table 2. For instance, both M1-2-1 and M2-1-2 involve an M1-2 and an M2-1 transition. In isolation, the mean best-mode  $Q_s$  for M1-2 is 0.226 and for M2-1 is 0.456; yet for M1-2-1 the mean best-mode  $Q_s$  is 0.509 and for M2-1-2 it is 0.497.

Finally, it is clear that in the single-schedule (no-shock) experiments of Table 1, 100% of the optimum  $Q_s$  values are non-CDA; while in the dual-schedule (single-shock) experiments of Table 2, 70% are non-CDA; and in the triple-schedule (dual-shock)

experiments of Tables 3 to 6, the proportion of non-CDA optima drops again to 56%. Thus, these new data add further weight to the conjecture (first made in [9]) that the more changes to the market supply and demand schedules during evaluation of a genotype, the more likely it is that the CDA  $Q_s=0.5$  value is the optimal mechanism. That is, *in the limit* when nothing is predictable in advance about the market supply and demand curves, the CDA is likely to be the optimal mechanism. A corollary to this is that if there is *some* regularity in the market supply and demand, then a hybrid auction mechanism *might* exhibit better dynamics than a CDA.

#### 4. Discussion

This paper extends the line of research first reported on in [6]. 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  $Q_s=1.0$ . Again, in the majority of our experiments, new “hybrid” market mechanisms were found to give better market dynamics than the previously-known auction styles. To reiterate: while such evolved market mechanisms are unlike any human-designed mechanism, they could nevertheless readily be implemented as online electronic marketplaces.

Thus, one contribution of this paper is the confirmation that the evolution of one-sided  $Q_s \sim 0.0$  results for M1 and M2 in [6] were 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 in the M1-2, M2-1, M1-2-1, and M2-1-2 cases,  $Q_s \sim 0.0$  is not the best value, even though it was the optimum for M1

and M2 individually. A second contribution is the confirmation that the optimum  $Q_s$  value is order-dependent in both the dual-schedule and the triple-schedule experiments: e.g. that the evolved value of  $Q_s$  for M1-2 is different to that for M2-1, and different again for M1-2-1 and M2-1-2. A third contribution is the hypothesis suggested by these data, i.e. that the CDA may be best when nothing can be predicted about the nature of the supply and demand curves, but that hybrid two-sided non-CDA mechanisms may be optimal when some regularity can be observed in the supply and demand schedules.

## 5. Related Work

The field of automated design of online auction markets by genetic algorithm is very new. To the best of my knowledge, it appears that the first paper in this field was the initial publication on evolving  $Q_s$  for ZIP-trader marketplaces [6]. The key results in that paper have since been replicated by Robinson [18] and by Qin [19]. In particular, Qin used a different genetic encoding that allowed true versions of the one-sided English and Dutch-Flower auctions to be evolved, but hybrid auction mechanisms were still settled on by the GA. Qualitatively similar results have also since been demonstrated in e-marketplaces populated by non-ZIP software-agent traders [10]. Results from a similar research project, using another evolutionary algorithm (i.e., genetic programming) for mechanism design in a different context, have subsequently been published [17]. Most recently, Bye [2] has published results from using a genetic algorithm to develop new forms of sealed-bid auction mechanism, independent of the intelligence (or lack of intelligence) of the traders taking part in those auctions. Significantly, Bye demonstrates that hybrid auction mechanisms (similar in spirit to the hybrid “non-standard”  $Q_s$  values evolved here) are found by the GA to be optimal for a number of realistic scenarios.



## 6. Conclusions

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 minority of the experiments reported here no such regularity was identified for exploitation, in the majority of our experiments there was an underlying regularity that allowed an evolved hybrid 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.* [11] 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 financial markets (populated by artificial trading agents) will be savings or profits measured in billions of dollars.

## Acknowledgements

Thanks to members of the HP Labs Biologically-Inspired Complex Adaptive Systems (BICAS) research group for valuable discussions. See [www.hpl.hp.com/research/bicas](http://www.hpl.hp.com/research/bicas).

## References

- [1] Bullock, S. (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.
- [2] Byde, A. (2002) "Applying Evolutionary Game Theory to Auction Mechanism Design". Accepted for presentation at the *2003 ACM Conference on E-Commerce*. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-321.
- [3] Cliff, D. (1997), "Minimal-intelligence agents for bargaining behaviours in market environments". Hewlett-Packard Laboratories Technical Report HPL-97-91.
- [4] Cliff, D. (1998), "Genetic optimization of adaptive trading agents for double-auction markets" in *Proceedings of Computational Intelligence in Financial Engineering (CIFE) 1998*. IEEE/IAFE/Informs (preprint proceedings), pp.252-258, 1998.
- [5] Cliff, D. (2001), "Evolutionary optimization of parameter sets for adaptive software-agent traders in continuous double-auction markets". Presented at the Artificial Societies and Computational Markets (ASCMA98) workshop at the Second International Conference on Autonomous Agents, Minneapolis/St. Paul, May 1998. Also available as HP Labs Technical Report HPL-2001-99.
- [6] Cliff, D. (2002), "Evolution of market mechanism through a continuous space of auction-types". Presented at *Computational Intelligence in Financial Engineering* session

at *CEC2002*, Hawaii, May 2002. Also available as Hewlett-Packard Laboratories Technical Report HPL-2001-326.

[7] Cliff, D. (2002), “Evolution of market mechanism through a continuous space of auction-types II: Two-sided auction mechanisms evolve in response to market shocks”. Presented at *Agents for Business Automation* session at *IC2002*, Las Vegas, June 2002. In: *Proceedings of the International Conference on Internet Computing IC02*, Volume III, edited by H.R. Arabnia and Y. Mun. CSREA Press, pp.682-688. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-128.

[8] Cliff, D. (2002) “Visualizing Search-Spaces for Evolved Hybrid Auction Mechanisms”. Presented at the *Beyond Fitness: Visualizing Evolution* workshop at the 8<sup>th</sup> International Conference on the Simulation and Synthesis of Living Systems (ALifeVIII) conference, Sydney, December 2002. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-291.

[9] Cliff, D. (2002) “Evolution of Market Mechanism Through a Continuous Space of Auction-Types III: Multiple Market Shocks Give Convergence Toward CDA”. Hewlett-Packard Laboratories Technical Report HPL-2002-312.

[10] Cliff, D., Walia, V., & Byde, A. (2002) “Evolved Hybrid Auction Mechanisms in Non-ZIP Trader Marketplaces”. Accepted for presentation at the *International Conference on Computational Intelligence for Financial Engineering (CIFEr03)*, Hong Kong, March 2003. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-247.

- [11] Das, R., Hanson, J., Kephart, J., & Tesauro, G. (2001), "Agent-human interactions in the continuous double auction" *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI-01)*, 2001.
- [12] Friedman, D., & Rust, J. (eds) (1993) *The Double Auction Market: Institution, Theories, and Evidence*. Addison Wesley.
- [13] Gode, D. & Sunder, S. (1993), "Allocative efficiency of markets with zero-intelligence traders" *Journal of Political Economy* **101**:119-137, 1993.
- [14] Goldberg, D. E. (1989) *Genetic Algorithms: In Search, Optimization, and Machine Learning*. Reading, MA: Addison Wesley.
- [15] Mitchell, M., (1998). *An Introduction to Genetic Algorithms*. MIT Press.
- [16] New York Stock Exchange (2002), *Stock Market Activity* report available at: [http://www.nyse.com/pdfs/02\\_STOCKMARKETACTIVITY.pdf](http://www.nyse.com/pdfs/02_STOCKMARKETACTIVITY.pdf)
- [17] Phelps, S., McBurney, P., Parsons, S., & Sklar, E., (2002) "Co-evolutionary auction mechanism design: a Preliminary Report", Technical Report ULCS-02-019, Department of Computer Science, University of Liverpool, 2002.
- [18] Qin, Z., (2002) *Evolving Marketplace Designs by Artificial Agents*. MSc Thesis, Computer Science Department, Bristol University, September 2002.
- [19] Robinson, N., (2002) *Evolutionary Optimization of Market-Based Control Systems for Resource Allocation in Compute Farms*. MSc Thesis, University of Sussex School of Cognitive and Computing Sciences, September 2002.
- [20] Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986) "Learning internal representations by error propagation", in *Parallel Distributed Processing: Explorations*

*in the Microstructures of Cognition*, Volume 1, edited by D. E. Rumelhart and J. L. McClelland (Cambridge, MA: MIT Press), pp. 318-362, 1986.

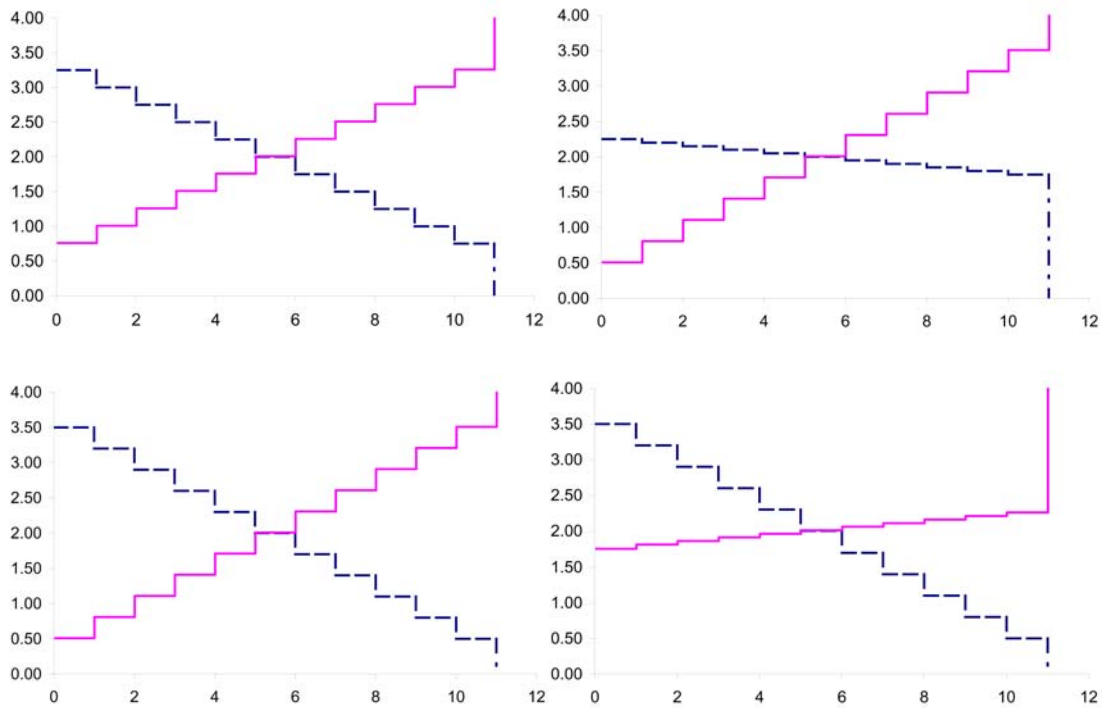
[21] Siegel, S., & Castellan, N. J., (1988), *Nonparametric Statistics for the Behavioral Sciences*. McGraw Hill.

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

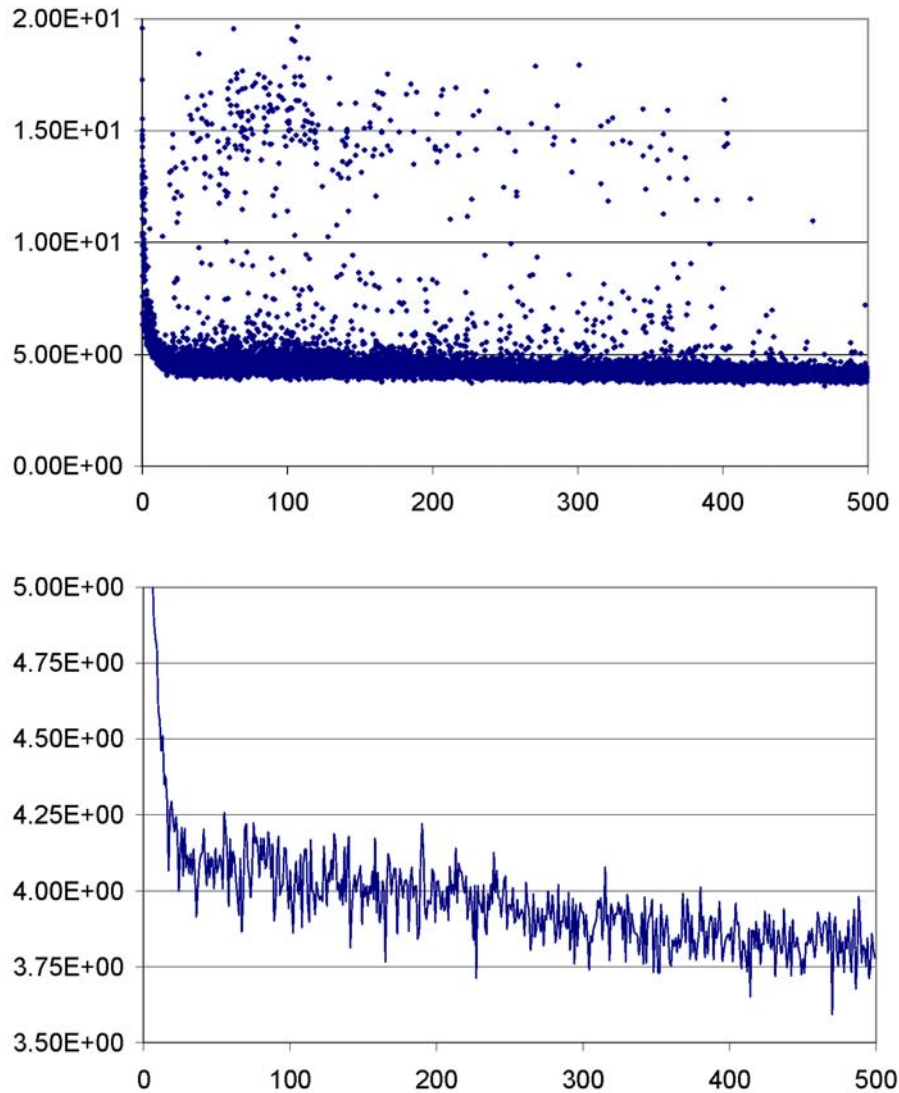
[23] Walia, V., (2002) *Evolving Market Design*, MSc Thesis, School of Computer Science, University of Birmingham, September 2002.

[24] Wilson, S.W., (1995) "Classifier Fitness Based on Accuracy" *Evolutionary Computation*, **3**(2):149-175, 1995.

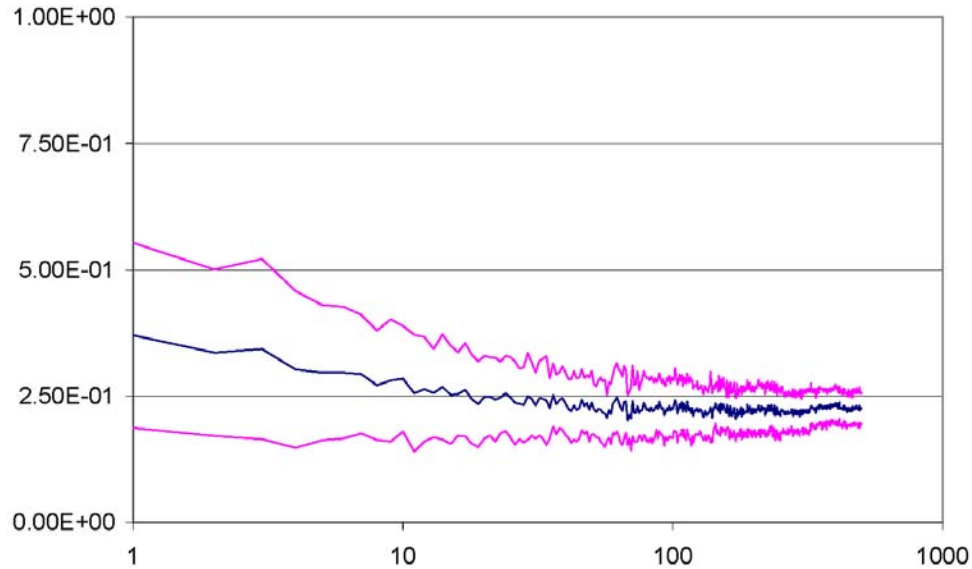
[25] Wurman, P.R., Wellman, M.P., & Walsh, W.E. (2001) "A Parameterization of the Auction Design Space" *Games and Economic Behavior* **35**:304-338, 2001.



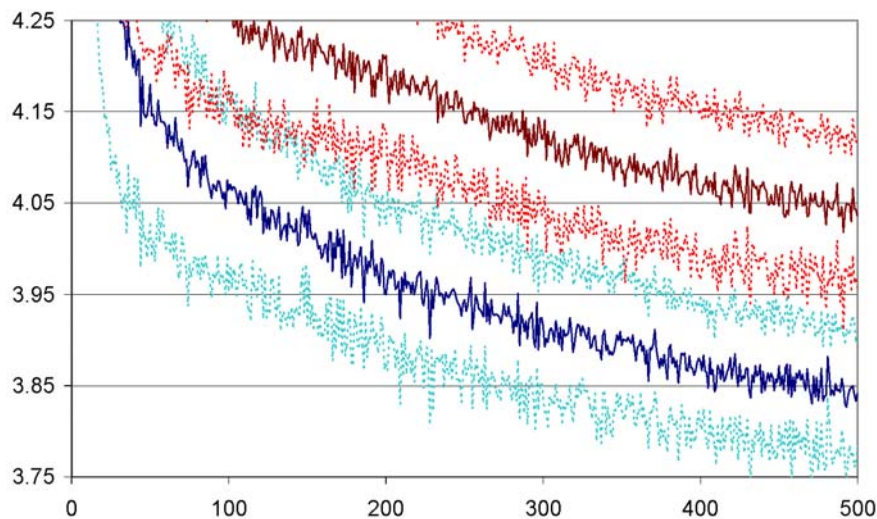
**Figure 1:** Supply and demand schedules for markets M1 (top left), M2 (top right), M3 (bottom left) and M4 (bottom right). In all three figures, the horizontal axis is quantity (from 0 to 12) and the vertical axis is price (from 0.00 to 4.00). The upward-sloping supply curve is shown by the solid line, and the downward-sloping demand curve is shown by the broken line.



**Figure 2a (upper graph):** fitness scores of all 30 members of the population for each generation. Horizontal axis is generation number (0 to 500); vertical axis is fitness score (0 to 20). **Figure 2b (lower graph):** Fitness score of the elite individual (i.e., the best genotype, with the lowest score) in each generation for the experiment shown in Figure 2a. Horizontal axis is generation number (0 to 500); vertical axis is fitness score (3.5 to 5.0).



**Figure 3:** Evolutionary trajectory of mean (plus and minus one s.d.;  $n=36$ ) value of  $Q_s$  in the best elite mode of the 50 experiments shown in Figure 3. Mean  $Q_s$  settles to  $\sim 0.23$



**Figure 4:** Average elite fitnesses from 50 EM and 50 FM( $Q_s=0.5$ ) MI-2 experiments; data is plotted for mean fitness, plus and minus one 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 ( $n=49$ ).



| Schedule  | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|-----------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| <i>M1</i> | 3.22      | 0.024        | 5    | 4.45      | 0.155        | 48   | Y   | 0.000                  | 0.0002                    |
| <i>M2</i> | 2.16      | 0.103        | 45   | 3.13      | 0.141        | 50   | Y   | 0.069                  | 0.0426                    |
| <i>M3</i> | 5.19      | 0.127        | 50   | 5.52      | 0.168        | 50   | Y   | 0.158                  | 0.0312                    |
| <i>M4</i> | 0.60      | 0.045        | 50   | 0.72      | 0.045        | 50   | Y   | 0.686                  | 0.0433                    |

**Table 1:** Summary of results from dual-schedule (single-shock) experiments. The column labelling is explained in the text.

| Schedule    | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|-------------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| <i>M1-2</i> | 3.85      | 0.058        | 36   | 4.04      | 0.078        | 49   | Y   | 0.226                  | 0.0309                    |
| <i>M2-1</i> | 4.18      | 0.102        | 46   | 4.18      | 0.092        | 50   | N   | 0.456                  | 0.0312                    |
| <i>M2-3</i> | 3.94      | 0.138        | 49   | 3.98      | 0.128        | 48   | N   | 0.561                  | 0.0264                    |
| <i>M3-2</i> | 3.05      | 0.056        | 49   | 3.46      | 0.082        | 50   | Y   | 0.137                  | 0.0254                    |
| <i>M1-4</i> | 2.78      | 0.061        | 36   | 3.08      | 0.069        | 50   | Y   | 0.211                  | 0.0263                    |
| <i>M4-1</i> | 2.79      | 0.094        | 50   | 2.97      | 0.093        | 50   | Y   | 0.380                  | 0.0237                    |
| <i>M4-3</i> | 3.01      | 0.131        | 50   | 3.25      | 0.118        | 50   | Y   | 0.364                  | 0.0184                    |
| <i>M3-4</i> | 3.17      | 0.078        | 50   | 3.47      | 0.083        | 50   | Y   | 0.212                  | 0.0294                    |
| <i>M2-4</i> | 3.57      | 0.128        | 49   | 3.59      | 0.117        | 49   | N   | 0.405                  | 0.0394                    |
| <i>M4-2</i> | 2.69      | 0.079        | 50   | 2.76      | 0.075        | 50   | Y   | 0.276                  | 0.0754                    |

**Table 2:** Summary of results from dual-schedule (single-shock) experiments. The column labelling and formatting is the same as for Table 1.

| Schedule | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|----------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| M1-2-1   | 4.35      | 0.084        | 46   | 4.32      | 0.076        | 49   | N   | 0.509                  | 0.0231                    |
| M2-1-2   | 3.92      | 0.073        | 50   | 3.91      | 0.076        | 49   | N   | 0.497                  | 0.0263                    |
| M2-3-2   | 2.99      | 0.075        | 49   | 3.00      | 0.097        | 49   | N   | 0.584                  | 0.0266                    |
| M3-2-3   | 3.87      | 0.070        | 50   | 3.86      | 0.087        | 50   | N   | 0.528                  | 0.0200                    |
| M1-2-3   | 4.24      | 0.066        | 50   | 4.28      | 0.076        | 50   | Y   | 0.564                  | 0.0238                    |
| M3-2-1   | 3.98      | 0.050        | 50   | 3.98      | 0.067        | 50   | N   | 0.473                  | 0.0218                    |

**Table 3:** Summary of results from dual-shock experiments where each shock involves a major change only to the demand curve. The column labelling and formatting is the same as for Table 1.

| Schedule | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|----------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| M1-4-1   | 3.25      | 0.083        | 40   | 3.75      | 0.083        | 50   | Y   | 0.187                  | 0.0225                    |
| M4-1-4   | 2.30      | 0.077        | 50   | 2.49      | 0.052        | 50   | Y   | 0.368                  | 0.0205                    |
| M4-3-4   | 2.83      | 0.056        | 50   | 2.85      | 0.057        | 50   | N   | 0.448                  | 0.0181                    |
| M3-4-3   | 3.52      | 0.083        | 50   | 4.21      | 0.083        | 50   | Y   | 0.146                  | 0.0213                    |
| M1-4-3   | 3.25      | 0.101        | 39   | 3.90      | 0.084        | 50   | Y   | 0.165                  | 0.0199                    |
| M3-4-1   | 3.56      | 0.082        | 49   | 4.07      | 0.086        | 50   | Y   | 0.173                  | 0.0230                    |

**Table 4:** Summary of results from dual-shock experiments where each shock involves a major change only to the supply curve. The column labelling and formatting is the same as for Table 1.

| Schedule | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|----------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| M4-3-2   | 2.25      | 0.096        | 50   | 2.53      | 0.091        | 50   | Y   | 0.348                  | 0.0226                    |
| M2-3-4   | 3.00      | 0.087        | 49   | 3.09      | 0.098        | 50   | Y   | 0.575                  | 0.0238                    |
| M4-1-2   | 2.97      | 0.078        | 50   | 3.11      | 0.067        | 50   | Y   | 0.379                  | 0.0188                    |
| M2-1-4   | 3.29      | 0.082        | 48   | 3.31      | 0.074        | 50   | N   | 0.492                  | 0.0300                    |

**Table 5:** Summary of results from dual-shock experiments where one shock involves a major change only to the demand curve and the other involves a major change only to the supply curve. The column labelling and formatting is the same as for Table 1.

| Schedule | EM: $\mu$ | EM: $\sigma$ | EM:n | FM: $\mu$ | FM: $\sigma$ | FM:n | 1%? | Q <sub>s</sub> : $\mu$ | Q <sub>s</sub> : $\sigma$ |
|----------|-----------|--------------|------|-----------|--------------|------|-----|------------------------|---------------------------|
| M2-4-2   | 3.83      | 0.088        | 50   | 3.95      | 0.096        | 50   | Y   | 0.332                  | 0.0276                    |
| M2-4-2   | 3.14      | 0.068        | 50   | 3.14      | 0.084        | 50   | N   | 0.496                  | 0.0271                    |

**Table 6:** Summary of results from dual-shock experiments where each shock involves major changes to both the supply curve and the demand curve. The column labelling and formatting is the same as for Table 1.