ZIP60: Further Explorations in the Evolutionary Design of Trader Agents and Online Auction-Market Mechanisms

Dave Cliff [1]

Foreign Exchange Complex Risk Group,
Deutsche Bank,
1 Great Winchester Street, London EC2N 2DB.

Email `dave.cliff@bcs.org`
Phone +44 (0)79 77 55 22 50

Abstract

The "ZIP" adaptive automated trading algorithm has been demonstrated to outperform human traders in experimental studies of continuous double auction (CDA) markets populated by mixtures of human and "software robot" traders. Previous papers have shown that values of the eight parameters governing behavior of ZIP traders can be automatically optimized using a genetic algorithm (GA), and that markets populated by GA-optimized traders perform better than those populated by ZIP traders with manually-set parameter values. This paper introduces a more sophisticated version of the ZIP algorithm, called "ZIP60", which requires the values of 60 parameters to be set correctly. ZIP60 is shown here to produce significantly better results in comparison to the original ZIP algorithm (called "ZIP8" hereafter) when a GA is used to search the 60-dimensional parameter space. It is also demonstrated here that this works best when the GA itself has control over the dimensionality of the search-space, allowing evolution to guide the expansion of the search-space up from 8 parameters to 60 via intermediate steps. Principal component analysis of the best evolved ZIP60 parameter-sets establishes that no ZIP8 solutions are embedded in the 60-dimensional space. Moreover, some of the results and analysis presented here cast doubt on previously-published ZIP8 results concerning the evolution of new 'hybrid' auction mechanisms that appeared to be improvements on the CDA: it now seems likely that those results were actually consequences of the relative lack of sophistication in the original ZIP8 algorithm, because "hybrid" mechanisms occur much less frequently when ZIP60s are used.

Keywords: Algorithmic Trading; Automated Market Mechanism Design; Adaptive Trader-Agents; ZIP Traders; Genetic Algorithms.

IEEE Trans. Evol. Comp. Special Issue on Evolutionary Computation for Finance and Economics
Themes: Agent-Based Computational Economics; Trading Strategies; Experimental Economics; Microeconomic Behavior.

[1] Author's current address: School of Electronics and Computer Science, The University of Southampton, Southampton SO17 1BJ, UK. E-mail: `dc@ecs.soton.ac.uk`.

1. Introduction and Background

1.1 Introduction

The Zero-Intelligence Plus ("ZIP") adaptive automated trading algorithm (Cliff, 1997) has been demonstrated to outperform human traders in experimental studies of continuous double auction (CDA) markets populated by mixtures of human and "robot" traders (Das, *et al.*, 2001). The CDA is an auction mechanism in which transaction prices observed in the marketplace reliably, rapidly, and robustly converge to the market's underlying theoretical competitive equilibrium price. Such *equilibration* behavior is highly desirable, and the CDA is used extensively in the world's financial markets. To successfully populate a CDA market with ZIP traders, the values of eight control parameters need to be set correctly. While these eight values can of course be set manually, previous papers have demonstrated that values of those parameters can be automatically optimized using a simple genetic algorithm (GA) to tailor ZIP traders to particular markets, producing results superior to those from ZIP traders with manually-set parameter values (Cliff, 1998, 2001). Furthermore, a simple extension of the GA-ZIP approach (i.e., adding a single additional real-valued numeric parameter) allows for mechanized market-mechanism design, and has been demonstrated as a possible way of automatically discovering novel forms of auction mechanism that appear to have better equilibration properties than the CDA (Cliff, 2002a-d, 2003). This paper introduces a more sophisticated version of the ZIP algorithm, which is known as "ZIP60", because it requires 60 real-valued control parameters to be set correctly, and thus the original ZIP algorithm is now in turn re-named "ZIP8". It is shown here that CDA markets populated with ZIP60 traders have significantly better equilibration behavior than markets populated with ZIP8 traders.

Manually identifying the correct values for 60 control parameters could be a very laborious task, but it is demonstrated here that an appropriate automatic optimization process can discover good sets of values for the parameters. The same simple GA as was used with ZIP8s, now operating in the 60-dimensional parameter space, produces ZIP60 traders with mean scores significantly improved over ZIP8s, but also with worryingly high variance in those improvements. A slight revision of the approach is shown to give results with even higher mean improvements and also with much lower variance in those improvements. The revised approach involves giving the GA control over the dimensionality of the parameter-space being searched so that the evolutionary search starts in an eight-dimensional space, and the GA then automatically and gradually expands the dimensionality of the search-space up to sixty-dimensional only when the increased number of parameters leads to identifiably better solutions.

Reliance on an offline ("batch-mode") optimizer such as the GA used here is something of an obstacle to real-world deployment of ZIP60. If we consider any one vector of 60 parameter values as a candidate "solution" for using ZIP60 in a real-world market or marketplace, then the GA can be thought of as evaluating a large number of candidate solutions against some number of test problems, with the intention that the best solution found by the GA could then be deployed in the real-world market. The problem with this approach is that the best solution will be well-tailored to the set of test-problems, and so if there is any mismatch between the test problems and the real market (or if the real market changes in some significant way while the GA optimizer is running) then the "best" solution found by the GA may perform poorly in the real market. Thus, the ultimate intention of this work is to develop "online" learning/adaptation techniques that can react to market conditions and events to adjust the ZIP60 parameters "on the fly", in real time. But, before that is done, the GA is used here to establish initial proof of the concept that there exist ZIP60 solutions that perform better than ZIP8s. Having established that existence proof in Section 3 of this paper, Section 4 goes on to analyze the evolved ZIP60 solutions, first to demonstrate that they do not involve ZIP8 solutions embedded in the 60-d space; and second in an attempt to identify any redundant dimensions in the ZIP60 parameter space that could be eliminated without significant change to the behavior of the ZIP60 traders. This search for redundancy is important because many online/real-time learning algorithms scale poorly as the number of dimensions increase. So, for instance, if it can be demonstrated that there is some set of 27 (say) basis vectors that define a subspace within the 60-d ZIP60 parameter space that eliminates all redundancy from the parameter set, then that is likely to make the successful application of an online learning method significantly more practicable

Furthermore, the results from ZIP60 presented in this paper, while better than ZIP8, show a markedly reduced incidence of cases where the GA discovers novel "hybrid" auction mechanisms, in which the ZIP traders perform significantly better than when they interact within the fixed CDA mechanism. A plausible interpretation of this is that it indicates that the earlier ZIP8 results (where improvements on the CDA were common) were actually consequences of the relative lack of sophistication in the ZIP8 algorithm, rather than consequences of previously-undiscovered weaknesses in the CDA mechanism that the ZIP8 traders were operating within.

So, in summary, the contributions of this paper are as follows:

- The ZIP60 algorithm, and a number of other higher-dimensioned-parameter-space extensions of the original ZIP8, are introduced.

- It is shown that difficulties encountered in searching a 60-dimensional space *ab initio* can be circumvented by commencing the search in a lower-dimensional space and then allowing the GA to automatically and gradually increase the dimensionality of the search space.

- Experimental results from the performance of GA-optimized ZIP60 traders are summarized and analyzed. These new results come from extensive simulation studies involving over 97 billion trading periods (simulated "days" in the experimental markets), and they empirically demonstrate the superiority of ZIP60 over ZIP8.

- The underlying GA-optimized parameter values are visualized and analyzed using multivariate statistics techniques, to explore whether any of the 60 parameters are unused or redundant. In all cases, the best solution involved more than 8 parameter values. Nevertheless, evidence is presented here that in some cases many fewer than 60 unique parameter-value need to be specified when using ZIP60, because the best values for some parameters may be dependent on (or derivable from) the values of one or more other parameters. For instance, the value of one parameter might be directly computable from a simple linear combination of the values of the other parameters, or perhaps as a nonlinear function of some other subset of parameters. This paper presents a proof-of-concept in analyzing the best 60-dimensional parameter-vectors found by the GA, identifying lower-dimensional subspaces in which the best solutions reside.[2]

- The results presented here cast some doubt on previously-published results, where a parameter governing the auction-market mechanism was also evolved (strictly, co-evolved) alongside the eight parameters governing the ZIP8 traders (Cliff 2002a-d, 2003). In those earlier results, it appeared that the GA was discovering new "hybrid" auction mechanisms that have better equilibration behavior than the CDA. The results presented here indicate that perhaps the earlier results are artifacts of populating the markets with ZIP8 traders: when ZIP60 traders are used in the same style of experiment, the evolved auction mechanisms differ from the CDA much less frequently.

In the interests of scientific openness and ease of replicability, the C source-code that was used to generate the results in this paper has been published as a technical report freely available on the web (Cliff, 2005).

This paper concludes at the point where the superiority of ZIP60 over ZIP8 has been demonstrated by using a GA optimizer to find appropriate settings of the sixty parameter-values. There are several open avenues of research that could be pursued to extend or further explore the ideas presented here. In particular, it is important to note that the results in this paper are certainly not intended as an absolute and conclusive demonstration that ZIP60 is superior to all other CDA bidding algorithms, or that the solutions discovered by the GA are optimal in the sense of the GA routinely discovering Nash equilibria in the experimental markets that ZIP60 is studied within here. Identification and analysis of Nash equilibria in the evolved solutions would require extensive additional empirical work, and anyhow the manifest limitations of the ZIP60 agents (in comparison to idealized omnipotent agents) mean that such equilibria may simply not be attainable: see, for example, (Gottlob, Greco, & Scarcello, 2005; Bowling & Veloso, 2004).

Moreover, this paper studies the equilibrating performance of markets that are homogeneously populated with one type of trader-agent, in the style of (Gode & Sunder, 1993; Cliff, 1997; Preist & van Tol, 1998; and Gjerstad & Dickhaut, 1998), rather than studying strategic interactions within markets heterogeneously populated by two or more different types of trading algorithms or market mechanisms, such as is exemplified by the work of (Tesauro & Das, 2001; Tesauro & Bredin, 2002; and Phelps, Parsons, & McBurney, 2004). The rationale for this paper's focus on studying interactions among populations of homogeneous traders exposed to dynamically changing supply and demand is straightforward. While the original paper (Cliff, 1997) that introduced the ZIP8 algorithm studied its performance only in homogeneously populated markets, ZIP8 was subsequently used as a benchmark trading algorithm in numerous studies of strategic interactions between heterogeneous mixes of trading algorithms, performed by several

[2] A full analysis of *all* the best ZIP60 parameter-set vectors found so far is way beyond the scope of this paper, but can be addressed in forthcoming publications.

independent groups of researchers. The number of such papers in which ZIP8 (or close derivatives of ZIP8) have been used is fairly substantial, and the list includes the following: (Das, Hanson, Kephart, & Tesauro, 2001; Tesauro & Das, 2001; Tesauro & Bredin, 2002; He, Leung, & Jennings, 2003; Phelps, Parsons, & McBurney, 2004; Vytelingum, Dash, David, & Jennings, 2004; and Bagnall & Toft, 2004). Thus, given that so much prior work exploring strategic interactions and heterogeneous populations has been based on ZIP8, it seems reasonable at least to presume that researchers with an interest in studying heterogeneous marketplaces might find ZIP60 a useful new benchmark, even though this current paper reports only on ZIP60 in homogeneous settings. The study of ZIP60's strategic interactions with other CDA bidding algorithms is certainly an important topic of further research, but it is not within the scope of this paper; this paper first seeks to establish that ZIP60 is a worthwhile replacement for ZIP8.

Furthermore, it is worth noting that in pretty much all of the above-cited papers studying strategic interactions between heterogeneous mixtures of bidding algorithms, the results come from experiments in which the nature of the market supply and demand curves are essentially fixed for the duration of each experiment. That is, with the notable exception of the paper by Das *et al.*, (2001), none of these studies explore the effects that sudden significant alterations to the supply or demand (or both) curves can have on the trader-agent performance and/or on market's internal dynamics: the supply and demand curves for the *initial* trading period in any one experiment are largely the same as the curves for the *final* trading period in that experiment. This seems very curious, considering that one commonly-claimed motivation for studying market systems is that mechanisms such as the CDA are interesting because of their ability to quickly and robustly adapt to dynamic and unexpected changes in supply and/or demand; and considering that studies of shock-changes in *human* CDA markets date back at least as far as Vernon Smith's seminal 1962 paper (Smith, 1962); and considering that such changes are widely known to occur in real-world markets.[3] This raises a significant question: if CDA markets are interesting because they exhibit attractive adaptation to dynamic changes in supply and demand, why is there this devotion in the published trading-agent literature to studying systems with essentially static supply and demand curves? In contrast, the results reported in this paper all come from experiments in which the marketplaces populated by ZIP60 traders periodically undergo sudden "shock" changes to the supply and/or demand curves, and where the ZIP60 traders are optimized on the basis of their ability to rapidly and stably adapt not only to the initial supply and demand schedule, but also to the new market supply and demand conditions prevailing after each shock-change.[4] So, in the experiments reported here, the final supply and demand curves in any one experiment may be significantly different from the initial ones. It seems reasonable to question whether the patterns of strategic interactions reported by the other authors cited above might possibly be different if they had used dynamically-changing rather than static supply and demand curves. For example: if strategy *S1* dominates strategy *S2* under static conditions, it may nevertheless be plausible that *S2* adapts better than *S1* in times of sudden significant change to supply and/or demand, so that if the two strategies are

[3] E.g., in high-frequency foreign-exchange price time series, "gap" step-changes in price are not unusual.

[4] The successful adaptive response of ZIP8 traders to shock-changes in the market's supply and demand curves was documented in the original 1997 paper that introduced ZIP (Cliff, 1997).

compared in a market with sufficiently turbulent supply and demand, *S2* could be found to dominate *S1*. Exploring this issue is another important topic for further research.

The rest of this paper is structured as follows. Section 1.2 gives some introductory and background material, which is necessarily very similar to previously-published accounts of the history and rationale for this work. Section 2 then 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 (e.g., Cliff 2002a; 2003), albeit extended to describe how the new experiments whose results are presented here differ from the previous work. Details of the extension to ZIP60, and results from experiments with ZIP60 traders, are presented in Section 3. The ZIP60 results are then analyzed and discussed in Section 4.

1.2. Background

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.[5] The CDA has been the subject of much study by economists, partially because it is so common (and hence 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 consistently taking place at off-equilibrium prices then someone somewhere in the market is being ripped off. Hence, rapid and stable 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. Friedman & Rust, 1993).

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

In the past three or four years, a leading-edge issue in the technical community who work within the world's financial markets has been the topic of *Algorithmic Trading*, where jobs performed traditionally by human traders are increasingly automated and executed by algorithms. More and more, algorithmic trading systems are becoming capable of going beyond mere replication of the roles of human traders, and they are now routinely capable of executing buy or sell orders in ways that are pretty-much impossible for a human trader, because the human is limited by being built from slow and noisy biological mate-

[5] For a review discussing the space of possible auction mechanisms, see (Wurman, Wellman, & Walsh, 2001); for an elementary introduction to issues in auction design, see (Rothkopf & Park, 2001).

rials. For example, a single big block trade (the sheer size of which might adversely affect the market price for the asset being traded) might be automatically broken up by the algorithmic trading system into a very large number of very small trades, which are then spread out over some time period to take advantage of trading opportunities as they occur, where each small trade is executed only a few microseconds after that trading opportunity opens up. The potential for ZIP traders to be used in algorithmic trading has caught the attention of journalists writing in the scientific press (Graham-Rowe, 2005), and in the financial press (Pritchard, 2005; Economist, 2005).

For the purposes of this paper, the world financial markets can be considered as just another set of instances of online e-marketplaces. This is obviously true for those centralized financial marketplaces (such as many major equity exchanges like the London Stock Exchange, or NASDAQ; and to a lesser extent the New York Stock Exchange) where there is a high degree of automation and machine processing of data. For those other major international financial systems where trading currently takes place in the absence of central exchanges (such as the $1900bn/day international foreign-exchange markets, where decentralized networks of broker-to-broker interactions still constitute much of the market), increasing penetration of appropriate automation technologies is clearly occurring, and so it is only a matter of time until they too are, in effect, online e-marketplaces.

ZIP (Zero-Intelligence-Plus) artificial trading agents, introduced in (Cliff, 1997), are software-agent “trader robots” that use a few heuristic rules and some simple machine learning techniques to adapt to operating as buyers or sellers in electronic versions of open-outcry auction-market environments. The market environments are similar to those used in Smith’s (1962) pioneering (and subsequently Nobel-prize-winning) experimental economics studies of the CDA and other auction mechanisms. ZIP traders were initially developed as a solution to the pathological failures of Gode & Sunder’s (1993) “ZI” (Zero-Intelligence) traders. The original motivation for inventing ZIP was to develop adaptive trading agents that could be used in distributed and decentralized market-based control systems (e.g., Clearwater, 1995), and also to extend the scope of research in adaptive behavior to include the study of interactions among economic agents within market mechanisms (Cliff & Bruten, 1999). However, a couple of years later, in some surprising empirical studies undertaken at IBM’s research labs by Das *et al.* (2001), it was shown that ZIP traders (and also IBM’s own “MGD” trading algorithm) consistently out-perform human traders in human-against-robot experimental-economics CDA marketplaces. In the IBM experiments, the robot traders consistently made profits a few percentage points higher than did the human traders they were competing against, and ZIP had the joint-highest average efficiency[6] of the algorithms used in the IBM study. In discussing the possible impact of their work, Das *et al.* (2001) wrote that the “…successful demonstration of machine superiority in the CDA … could have a … powerful financial

[6] Mean values calculated from the efficiency data presented in Table 1 of (Das *et al.* 2001) are 1.030 for ZIP (mean from n=2 experiments); 1.023 for MGD (n=4); and 0.876 for humans (n=6). Higher efficiency values are more desirable. The difference between the mean values for ZIP and MGD is so small that it seems highly unlikely to be statistically significant. Efficiency is defined as follows. For each trader i, let s_i denote the trader’s *actual surplus*, defined as the absolute difference between the transaction price and that trader’s limit price for the transaction. Let s_i* denote the trader’s *theoretical surplus* that would result if the transaction had taken place at the market’s theoretical competitive equilibrium price P_0.. Efficiency is the ratio of s_i to s_i*.

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 fact that ZIP did no worse than MGD in the heterogeneous human-robot studies reported by Das *et al.* (2001) is curious, and something of an anomaly. Previous studies by the IBM team had demonstrated that in similar studies where *all* the traders were robots, MGD did better than ZIP (Tesauro & Das, 2001)**;** and in a paper published after the Das *et al.* (2001) human-robot interaction studies, Tesauro & Bredin (2002) presented details of an algorithm called GDX which outperforms both MGD and ZIP in robot-vs-robot studies. GDX is based on MGD, extended to incorporate techniques from Dynamic Programming. Thus, while MGD and GDX have both been shown to outperform ZIP when there are *no* human traders active in the market (with GDX being dominant), the only presently available rigorous experimental results from human-robot interactions show ZIP and MGD as performing *equally* well, and *both* outperforming humans. Understanding why this intransitivity exists is clearly a topic for further research.

The style of auction used in the experiments reported here follows that of the original ZIP studies (Cliff, 1997) which themselves followed Smith's (1962) lead in having the traders react only to the most recent quote in the marketplace. That is, the studies here maintain no persistent *order book* showing the price-ordered list of best offers, and the price-ordered list of best bids, currently being quoted by the traders active in the market. In this sense, the experimental markets studied here are closer to the real-world markets for foreign exchange (FX) than they are for equities: most prominent national/international equity exchanges operate on an order-book basis; whereas the international FX markets do not yet do so, even where screen-based transactions account for much of the trading volumes (however, a transition to showing order books or "price ladders" on FX dealing screens currently appears to be finally getting underway). Adapting ZIP to operate in order-book-based markets is relatively trivial. Das *et al.* (2001), and Tesauro & Das (2001), each used ZIP in dynamic asynchronous order-book-based marketplaces, and they report that for their studies of strategic interactions they independently arrived at the same relatively minor extensions of ZIP that had earlier been developed by Preist & van Tol (1998). Certainly the extensions required of ZIP were sufficiently minor that the authors of the IBM papers did not see a need to re-name ZIP, whereas their altered and extended versions of the original GD algorithm (Gjerstad & Dickhaut, 1998) were re-named MGD (Tesauro & Das, 2001; Das *et al.*, 2001) and then GDX (Tesauro & Bredin, 2002).

The operation of ZIP traders has been successfully demonstrated in experimental versions of CDA markets similar to those found in the international financial markets, and in posted-offer auction markets similar to those seen in domestic high-street retail outlets (Cliff, 1997); and ZIP has also recently been adapted for use in sealed-bid auctions (Bagnall & Toft, 2004; 2005). There are a number of numeric parameters that govern the adaptation and trading processes of ZIP traders, and these parameters need somehow to be assigned values if a marketplace is to be populated by ZIP traders. In the original 1997 version of the ZIP algorithm, the parameter-values were set by hand, using "educated guesses". However, subsequent papers (Cliff, 1998, 2001) presented the first results from using a standard evolutionary computation technique – i.e., a simple "plain vanilla" genetic algorithm (GA) – to automatically optimize these parameter values, thereby eliminating the need for skilled human input in deciding the values. The GA found parameter-settings that were clear improvements on those chosen by hand by the inventor of the algorithm. All experience thus far with ZIP-trader markets has indicated that the market dynamics are not significantly affected as the size of the market (i.e., the number of traders active in the market) increases.

The present paper continues the exploration (documented in Cliff, 2002a-d, 2003) of some specific consequences of asking the following question: if, as Das *et al.* (2001) speculate, trader agents will come to replace human traders at the point of execution 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 continues exploring the application of one type of automated search/optimization algorithm, namely the GA (Goldberg, 1989; Mitchell, 1998), but any of many other types of automated search or optimization process could have been used instead of the simple GA used here (further details of which are given in Appendix C).

Prior to the research described in (Cliff, 2002a), in all previous work using artificial trading agents – ZIP or otherwise – the market mechanism (i.e., the type of auction the agents are interacting within) had been fixed in advance. Well-known market mechanisms from human economic affairs include: the English Auction (where sellers stay silent and buyers quote increasing bid-prices); the Dutch Flower Auction (where buyers stay silent and sellers quote decreasing offer-prices); the Vickery or second-price sealed-bid auction (where sealed bids are submitted by buyers, and the highest bidder is allowed to buy, but at the price of the *second-highest* bid); and the CDA (where sellers can announce decreasing offer prices while *simultaneously and asynchronously* the buyers can 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, all in the absence of an auctioneer). 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 (Cliff, 2002a). 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 was nothing to prevent the GA from settling on solutions that correspond to the known CDA mechanism type or the two entirely one-sided mechanisms, it was found that "hybrid" mechanisms, partway between the CDA and the one-sided mechanisms, could lead to the most desirable market dynamics. The significance of this is that 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 (Cliff, 2002a) were the first demonstration that radically new market mechanisms for artificial traders may be designed by automatic means. At much the same time, Steve Phelps and his colleagues were independently working on a conceptually very similar (but algorithmically rather different) theme of using artificial evolution to develop and study new auction-market mechanisms (Phelps *et al.* 2002; Phelps, Parsons, & McBurney, 2004; Phelps *et al.* 2005). Specifically, the methods used by Phelps *et al.* were initially genetic programming (see e.g. Koza, 1992; Koza *et al.*, 1998) and latterly replicator dynamics studies (see e.g. Taylor & Jonker, 1978; Witt, 1992; Borgers & Sarin, 1997)

Although the performance of the evolved market mechanisms is typically only better by a few percentage points (or even only a few basis points – i.e. a few hundredths of a percentage point) than that of the established human-designed mechanisms, the economic consequences could nevertheless 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 computer-designed 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 those operated by LSE, Euronext, and Deutsche Börse, and at similar exchanges elsewhere around the globe. And they're just the equity exchanges: the decentralized, global, 24x7, foreign exchange (FX) market is (at approximately $1900bn/day) a much more liquid market that offers proportionately higher potential gains.

While the results presented here turn out to cast some doubt on the earlier results from evolving hybrid market mechanisms with ZIP traders, it is important to note that those doubts are not actually a conclusive demolition. Moreover, the improved versions of ZIP introduced in this paper are perfectly capable of operating in regular CDA auctions.

The text that follows in Section 2 presents a summary of the methods used in the past work on GA-optimization of ZIP traders and of the market mechanisms that they operate within (Sections 2.1 and 2.2); it then summarises the previous results (Section 2.3); discusses related work by other researchers (Section 2.4); and ends with a critique and new analysis of the old results (Section 2.5).

2. Methods

2.1 Eight-parameter Zero-Intelligence-Plus (ZIP8) Traders

The original eight-parameter ZIP trading agents were described fully in a lengthy report (Cliff, 1997), which included sample source-code in the C programming language. For the purposes of the current discussion a high-level description of the algorithm and its eight key parameters is sufficient, although additional details for the algorithmically curious are given in Appendix B of this paper. Illustrative C source-code for the ZIP60 extension has been published in (Cliff, 2005). As will be seen in Section 3, there are in fact a family of ZIP algorithms between ZIP8 and ZIP60, and so hereinafter the acronym "ZIP" with no numeric suffix is intended to mean "all ZIPn for $8=n=60$, and beyond".

ZIP traders deal in arbitrary abstract commodities. Each ZIP trader i is given a private (i.e., secret) limit-price, l_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 l_i price then it generates zero utility ("profit" for the sellers or "saving" for the buyers), but utility greater than zero is preferred. For this reason, each ZIP trader i maintains a time-varying utility margin $m_i(t)$ and generates quote-prices $p_i(t)$ at time t according to $p_i(t)=l_i(1+m_i(t))$ for sellers and $p_i(t)=l_i(1-m_i(t))$ 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 l_i value. Trader i is given an initial value $m_i(0)$ (i.e., $m_i(t)$ 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 (Rumelhart, Hinton, & Williams, 1986) and in learning classifier systems (Wilson, 1995). This rule has a "learning rate" parameter b_i that governs the speed of convergence between trader i's quoted price $p_i(t)$ and the trader's idealized "target" price $t_i(t)$. When calculating $t_i(t)$, ZIP traders introduce a small random *absolute* perturbation with magnitude generated from[7] $\mathbf{U}[0,c_a]$ (this perturbation is positive when increasing $t_i(t)$, negative when decreasing) and also a small random *relative* perturbation generated from $\mathbf{U}[1-c_r,1]$ (when decreasing $t_i(t)$) or $\mathbf{U}[1,1+c_r]$ (when increasing $t_i(t)$), where c_a and c_r are global system constants. To smooth over noise in the learning system, there is an additional "momentum" parameter g_i for each trader (such momentum is also commonly used in back-propagation neural networks).

Thus, adaptation in each ZIP trader i has the following parameters: initial margin $m_i(0)$; learning rate b_i; and momentum term g_i. In an entire market populated by ZIP traders, values for these three parameters are randomly assigned to each trader via the following expressions: $m_i(0)=\mathbf{U}[m_{min}, m_{min}+m_D]$; $b_i=\mathbf{U}[b_{min}, b_{min}+b_D]$; and $g_i=\mathbf{U}[g_{min}, g_{min}+g_D]$. Note that any parameter p assigned a value from a uniform distribution $p=\mathbf{U}[a,a+b]$ can (if necessary) be implicitly assigned as a constant c by setting $a=c$ and $b=0$, *i.e.* $p=\mathbf{U}[c,c]$. So, defining the assignment of parameter values as being from random distributions includes the possibility that any or all of the parameters are given fixed/constant values.

Hence, to initialize an entire ZIP-trader market it is necessary to specify values for the six market-initialization parameters m_{min}, m_D, b_{min}, b_D, g_{min}, and g_D. Values also need to be specified 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. For the purposes of the narrative in this paper, we can consider the GA optimizer as a "black box" and leave it largely un-discussed: some details are given in Appendix C of this paper, and full details accompany the source-code in (Cliff, 2005).

[7] Here $v=\mathbf{U}[x,y]$ denotes a random real value v generated from a uniform distribution over the range $[x,y]$.

In the experiments reported here, as before, the passage of time is simulated by dividing continuous time (possibly irregularly) into discrete *slices*, numbered sequentially, where one significant event is known to occur in each slice. 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. (In contrast, Das *et al.* (2001) used a continuous-time formulation of the ZIP algorithm).

In the markets described here (as in Cliff, 1997, 1998, 2001, 2002a-d, 2003, 2005), 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 $m_i(t)$ 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 l_i limit prices, until either a maximum number of time-slices have run, or a maximum number of sequential quotes have been ignored.

Appendix B shows pseudo-code and gives further details of the update equations for the ZIP trading algorithm.

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 an unbiased 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 all previous ZIP work, that random binary variable is always independently and identically distributed (i.i.d.) 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. For the CDA we can then write $Pr(Q=s)=0.5$. 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. The fact that any ZIP trader j will accept a quote whenever $q(t)$ and $p_j(t)$ cross means that the one-sided extreme cases $Q_s=0.0$ and $Q_s=1.0$ are not *exact* analogues of the EA and DFA. Nevertheless, the methods developed in (Cliff, 2002a) have since been independently replicated and extended to include true analogues of the EA and DFA (Qin, 2002; Qin & Kovacs, 2004).

Here, as in (Cliff, 2002a), the Q_s values of 0.0, 0.5, and 1.0 are not considered 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, as a result of an *exogenously imposed* bias. While the history of human economic affairs seems to offer no examples of markets in which values of Q_s other than 0.0, 0.5, or 1.0 have been used. 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 (Cliff, 2002a-d, 2003) a ninth dimension was added to the search space, so the genotype in the GA became the eight real values for ZIP-trader initialization, plus a real value for Q_s. For completeness, note that (as in all the previous GA-ZIP work) in the new experiments reported here no “NYSE” quote-improvement or spread-reduction rule was used. Such a rule forces buyers to only quote bids higher than the current best bid, and sellers to only quote offers lower than the current best offer.

Finally, it is important to note that while the model of the CDA provided here by setting $Q_s=0.5$ imposes no *exogenous* bias on the quoting activity of the traders in the marketplace, it does not address the issue of spontaneous *endogenous* biases that might arise in a real market. A hypothetical example of endogenous biases might be a case where an erroneous news story causes a temporary crash in the market price of a security: as the price crashes downwards, there may be a simple absence of any bids from buyers for some period until the offer-prices quoted by the sellers have dropped far enough (so the market acts for a period as if $Q_s=1.0$); but then, once it is clear that the news story was wrong, as the market reverts back toward the previous “true” price, there is likely to be much more quoting activity from the buyers than from the sellers (so the market may appear for a period to be acting as if $Q_s=0.0$). As in all the previously-published GA-ZIP work, the basic ZIP-trader algorithm used here does not involve spontaneous generation of quotes by the traders (rather, any ZIP trader provides a quote when it is randomly selected to do so), so the study of such endogenously-arising biases is not addressed in the work reported here..

2.3. Previous ZIP8 Results

In (Cliff, 2003), results from 32 sets of ZIP8 experiments were published, where each experiment involved 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. For consistency and ease of comparison with those earlier results, all the new ZIP60 experiments whose results are tabulated in Section 3 of this paper will use the same set of schedules, and much the same experiment methods.

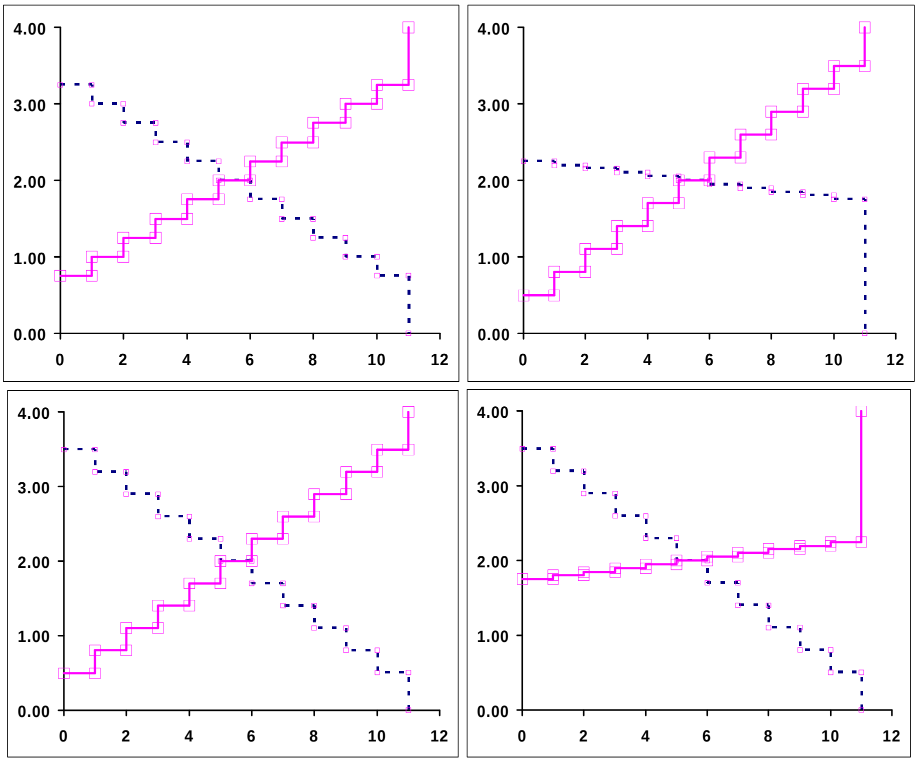

Figure 1: *Supply and demand schedules for markets M1 (top left), M2 (top right), M3 (bottom left) and M4 (bottom right). In all four figures, the horizontal axis is quantity (from 0 to 12) and the vertical axis is price (from 0.00 to 4.00). The upward-sloping supply curve is shown by the solid line,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

In all four schedules there are 11 buyers and 11 sellers, each empowered to buy/sell one unit of commodity: these relatively small numbers of traders are the cause of the stepped supply and demand curves. So, for example, in Market M1 the Supply curve looks the way it does because there is one seller willing to sell one unit at a price of $0.75; there is another seller willing to sell one unit at a price of $1.00 (so total supply at price $1.00 is 2 units); another one unit offered at a price of $1.25 (taking total supply to 3); and so on in price-increments of $0.25 up to the final (11th) unit offered at $3.25. The Demand curve is similarly the aggregate result of each trader having a cash endowment to buy one unit at a price no greater than that trader's limit price. Market M1 is taken from Smith's seminal 1962 paper on his early experimental economics work, and 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 (Cliff, 2002c) demonstrated that the minor differences between the supply and demand curves in M1 and M3 can lead to significant differences in the final best evolved solutions.

The evolving-mechanism experiments reported in (Cliff, 2003) studied the effects of "shock changes" being inflicted on the market by swapping from one schedule to another partway through the evaluation process: a procedure common in human-based experimental economics (see Smith (1962)), and something that ZIP8 was demonstrated as dealing robustly with in (Cliff, 1997). The maximum number of shocks occurring during any evaluation process in (Cliff, 2003) was two (i.e., switching between three schedules).[8] For instance, in one experiment referred to here as M121, the evaluation involved six trading periods ("days") with supply and demand determined by M1, then a sudden change to M2, then six periods/days later a reversion to M1 for a final six periods. The other sets of experiments are similarly named M212, M123, M321, and so on; and in all of them each of the three market schedules was used for six "days", so these two-shock trials all last for a total of 18 days.

As in the previous GA-ZIP work, the evaluation function was a weighted average of Smith's (1962) "α" measure of root mean square deviation of transaction prices from the underlying theoretical equilibrium price at the start of the period, measured across the six periods for each schedule used. Specifically, in each trading period p the value α_p was calculated, and the evaluation score was computed as $(1/\Sigma w_p).\Sigma(\alpha_p.w_p)$ for p=1…18 with weights w_1=1.75, w_2=1.5, w_3=1.25, $w_{3<p<7}$=1.0, $w_{p>6}=w_{p-6}$ for p>6.

In any one experiment, here involving a population of 30 genotypes over 500 generations, in each generation the elite (best-scoring) individual is of most interest, and so the time-series trajectory of the elite evaluation score for the population is compiled across the 500 generations. These results are non-deterministic: different runs of the GA (with different seed values for its random number generator) will yield different elite trajectories. Examining the elite-score trajectories from 50 repetitions of an experiment (varying only the random seed between repetitions) often gives multimodal results, and in all experiments we are interested only in the best elite mode (i.e. the mode with lowest scores), which could be summarized by the mean and standard deviation (s.d.) of the scores within that mode at each generation. However, relying on manual identification of the elite mode (as was done in (Cliff, 2003)) is subjective and labor-intensive, and automatically identifying the elite mode via cluster analysis does not entirely eliminate subjectivity either (Jain, Murty, & Flynn, 1999), so in this paper we instead simply study the results from the best 10% (i.e., the upper decile) of the 50 repetitions of each experiment.

Experiments where the GA controlled the value of Q_s are referred to here as evolving-mechanism (EM) conditions; and those where the value of Q_s was *not* evolved but instead was fixed at the CDA value of Q_s=0.5 are referred to here as fixed-mechanism (FM).

The results from 18 dual-shock (triple-schedule) experiments were presented in four separate data-tables in (Cliff, 2003), grouped by the nature of the shocks (i.e., the "treatment regime"). Table 3 showed results from experiments where only the demand curve undergoes a major change on each shock (i.e.: M121, M212, M232, M323, M123, and

[8] Shipp's (2004) MSc thesis explored the effects of extending the number of market shocks in the sequence.

M321). Table 4 showed results from experiments where only the supply curve undergoes a major change on each shock (i.e.: M141, M414, M434, M343, M143, and M341). In Table 5, 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 (i.e.: M432, M234, M412, and M214); and in Table 6 each shock involved a major change to both the supply curve and the demand curve (i.e.: M242 and M424). In this paper, all 18 dual-shock results are shown together in single graphs, but the results appear in table order, as was just listed. For convenience of later discussion, we will refer to these four groupings by the abbreviations T3, T4, T5, and T6, i.e. referring to each regime type by the table-numbers in (Cliff, 2003). So, to be clear:

- T3: two successive shock-changes to the Demand.
- T4: two successive shock-changes to the Supply.
- T5: one shock change to Demand, the other to Supply (in either order).
- T6: each shock involves major changes both to Supply and to Demand.

Analysis of the ZIP8 results from all four of these regimes, published in (Cliff, 2003) showed that the GA never failed to discover EM genotypes that were at least as good (i.e. had elite evaluation scores at least as low) as the corresponding FM genotypes, and in several cases the EM result was statistically significantly better (i.e., lower) than the FM result, at the 1% confidence level using the Wilcoxon-Mann-Whitney (see, e.g., Siegel & Castellan, 1988). This happened in one out of the six T3 treatments; in five of the six T4 treatments; in three of the four T5 treatments; and in one of the two T6 treatments. Shortly after the analysis of ZIP8 results in (Cliff, 2003) was published, a paper by Feltovich (2003) exposed serious problems with the Wilcoxon-Mann-Whitney (WMW) test, advocating the Robust Rank-Order (RRO) test as a replacement (see also Feltovich, 2006). The results of re-testing the significance of the ZIP8 EM/FM data in (Cliff, 2003), using RRO at 1% confidence in place of WMW, came out differently: now three of the six T3 treatments; and all six of the T4 treatments, showed a significant difference (T5 and T6 were unchanged). Prima facie, it would appear that the outcome of the RRO tests strengthens the results, because a greater number of significant differences are identified. Unfortunately, it’s not that simple: for the T3 sequence M121, the RRO test reveals a significant difference in the ZIP8 in favour of FM rather than EM (Cliff, 2005).

The histogram in Figure 2 shows the results for GA-optimized ZIP8 in FM and EM conditions. The statistics in Figure 2 are calculated from the same data as was originally analyzed and tabulated in (Cliff, 2003), but the statistics in Fig 1a are the results of conducting a more rigorous and careful analysis, the rationale for which is discussed in (Cliff, 2005). In the new analysis, the final score recorded as the outcome of any one experiment is now taken as an average of the final ten elite evaluation scores (specifically, over generations 490 to 500) to smooth over noise in the evaluation process; and the summary statistics for each type of experiment are here always calculated from the top 10% (i.e., the upper decile) of the 50 repetitions of each type of experiment, regardless of how many repetitions converged on solutions with final elite scores in the best elite mode. So, the data in Figure 2 show the mean and s.d. of the final-outcome elite scores from the best (lowest-scoring) five experiments in each study, and the RRO test is used at the 1% confidence level to detect significant differences between the EM and FM results.

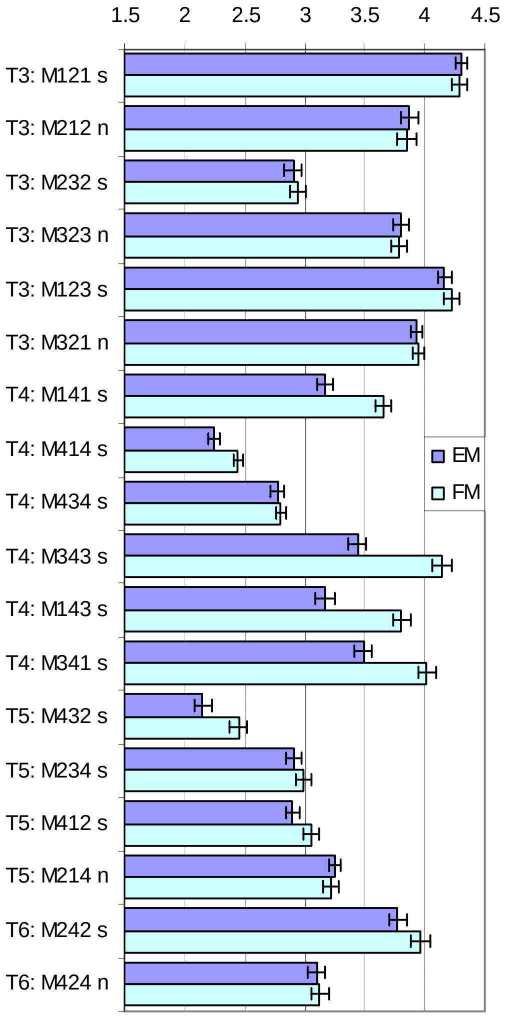

Figure 2. *ZIP8 fitness scores from tests in each of 18 market shock sequences, grouped into treatment types T3, T4, T5, and T6. Horizontal axis is evaluation score (a weighted average of root mean square deviation of the transaction price from the theoretical equilibrium price, expressed as a percentage of the equilibrium price); lower scores are better. The dark-shaded bars are the evolved-mechanism (EM) data; the light-shaded bars are fixed-mechanism (FM) data from using the CDA model. Error bars indicate the range plus and minus one standard deviation of the final results for the best n=5 experiments from i.i.d. 50 repetitions of each experiment. Each experiment's final result is calculated as the mean elite score over the last ten generations. The suffix "s" or "n" on each bar label indicates whether a statistically significant difference is detected ("s") or not ("n") between the EM and FM results, using the Robust Rank-Order (RRO) test at the 1% confidence level.*

2.4 Related Work

These previous GA-ZIP results have subsequently been replicated, adapted, and extended in a number of independent studies. Robinson (2002) explored the use of evolved market-mechanisms in the context of market-based control (e.g. Clearwater, 1995) of scarce resources in utility-scale corporate data centers. Walia *et al.* (Walia 2002; Cliff, Walia, & Byde, 2003) explored the use of the same evolving-mechanism techniques but with markets populated by Gode & Sunder's (1993) ZI trader-agents rather than ZIP traders, again finding evidence that non-standard hybrid mechanisms were discovered as good/best solutions by the GA; and Byde (2003) demonstrated that the same techniques could lead to the evolution of hybrid sealed-bid auction mechanisms, regardless of the type of trader operating in the market. Shipp (2004) investigated how the nature of the evolved solutions changed as the number of "market shocks" used in the evaluation process increased; and Wichett (2004) explored a system in which multiple reproductively separate "gene-pools" of ZIP traders competed, co-adapted, and co-evolved along with the market mechanism. Other recent uses of ZIP include modifying it for bargaining in sealed-bid auctions (Bagnall & Toft, 2004); using ZIP traders to study speculative trading in business-to-business exchanges (Li & Smith, 2004); and using ZIP traders to explore issues of reputation and information quality in a variety of market configurations (Ladley & Bullock , 2005)

In addition to the work of Phelps *et al.* discussed in Section 1.2 (Phelps *et al.*, 2002; Phelps, Parsons, & McBurney, 2004; Phelps *et al.*, 2005), which was undertaken at approximately the same time as the research summarized in Section 2.3, a number of other authors have more recently reported on the results of using artificial evolution and other forms of automated search, learning, or optimization for exploring spaces of possible trader-agent strategies, and possible new auction mechanisms, generally with positive results (Tesauro & Bredin, 2002; Gerding, Somefun, & La Poutré, 2004; Lochner & Wellman, 2004; Park, Durfee, & Birmingham, 2004; Greenwald *et al.*, 2005; Pardoe & Stone, 2005; Reeves *et al.*, 2005; Wellman, *et al.*, 2005). Of course, the paper introducing ZIP (Cliff, 1997) was not the first-ever study of artificial trading agents in double-auction markets; relevant prior work includes that by Wilson (1987), Friedman (1991), Easley & Ledyard (1992), Rust, Miller, Palmer (1992), and Gjerstad & Dickhaut (1998). For additional discussion of previous work, see the review in (Cliff, 1997).

In the next Section, the expansion of the number of parameters governing ZIP from 8 up to 60 is explained (Section 3.1), and empirical results are presented (Section 3.2) which show that ZIP60 performs better than ZIP8, so long as the GA is given control of the number of parameters being used by the ZIP algorithm, as described in Section 3.3. The results are then analysed and discussed in Section 4.

3. ZIP60

3.1 From 8 to 60 in five paragraphs

Despite the problem with the old M121 results that was revealed by the new analysis in the previous section, the results from using a GA to fine-tune the ZIP trader parameters remain generally encouraging. In fact, they are sufficiently encouraging to prompt speculation that perhaps new variants of ZIP can be developed to take advantage of the fact that we can now (generally, at least) rely on automated search and optimization algorithms such as the GA to set appropriate values for the numeric parameters affecting the ZIP market, so there is no need to try to keep the number of such parameters sufficiently small to render them easily manageable or comprehensible by a human.

To this end, note that in ZIP8 the genome specifies the same vector of eight real values $\{m_{min}, m_D, b_{min}, b_D, g_{min}, g_D, c_a, c_r\}$ whether the trader is a buyer or a seller. But in some situations it's perfectly plausible that the overall market dynamics might be better if the buyers were using different parameter-values to the sellers, so we could in principle have a GA-ZIP system dealing with these two cases (i.e. where *Case 1* is that the trader is a buyer; *Case 2* is that the trader is a seller) and hence optimizing sixteen real parameters (i.e., “ZIP16”), where the first eight values are the vector used to initialise the buyers and the second eight are the vector used to initialise the sellers.

Furthermore, note also that in some situations a ZIP trader (whether it is a buyer or a seller) has to increase its margin, and in others it has to decrease its margin, and that it might be useful to have different parameter-values depending on which of the four cases we are in, i.e. whether the trader is a buyer raising its margin, a buyer lowering its margin, a seller raising, or a seller lowering. That would give us four cases, each with eight values, or “ZIP32”.

But we can then note that, in the original specification of the ZIP algorithm, both for buyers and for sellers, there are actually *three* different cases or circumstances in which the trader alters its margin (see Cliff, 1997, pp.42-43 for the reasoning behind this design). For example, a seller's margin is *raised* if *one* condition holds true (i.e., if the last quote was accepted and the seller's current price is less than the price of the current quote); but a seller's margin is *lowered* if either of *two* other possible conditions are true (i.e., if the last quote was an accepted bid and the seller is active and the seller's price is greater than the price of the last quote; *or* if the last quote was an offer that was accepted and the seller is active and its price is greater than the price of the last quote). So we could have the genome specify *three* corresponding parameter-value vectors for the buyers and also *three* such vectors for the sellers, i.e. a total of six different vectors for six different cases, which at eight values per vector gives us “ZIP48”.

And in a final flourish of parameter-count inflation, let's abandon the use of a mere pair of system-wide global constants c_a and c_r and in place initialise each trader *i* with its own corresponding “personal” values $c_{a,i}$ and $c_{r,i,}$ generated at initialization from the uniform distributions $\boldsymbol{U}[c_{a:min}, c_{a:min}+c_{a:D}]$ and $\boldsymbol{U}[c_{r:min}, c_{r:min}+c_{r:D}]$. This addition of extra parameters still allows solutions involving the old system-wide constant c_a and c_r values to be “discovered” by the GA – that will happen if increased fitness values are associated with (near-)zero values of $c_{a:D}$ and $c_{r:D}$. So, each of the six parameter-value vectors needs now to specify values not only the six previous system parameters (m_{min} , m_D , b_{min} , b_D , g_{min} , and g_D) but also the values for the four newly-introduced system parameters $c_{a:min}$, $c_{a:D}$, $c_{r:min}$, and $c_{r:D}$ – for six cases, that gives six vectors, each with ten values per vector, and hence sixty values for ZIP60.

It is worth noting that this final increase from eight to ten parameter-values per case could also be applied to any of the other ZIP*n* versions mentioned in the preceding paragraphs. That is, by the expansion of the specification of c_a and c_r, ZIP8 could be expanded to ZIP10; ZIP16 could be expanded to ZIP20; and ZIP32 could be expanded to ZIP40.

The analysis and discussion that come later will be eased if we introduce some terminology here. While a ZIP8 trader has one genetically-specified value for each parameter (so, for example, it has only one b_{min} value), a ZIP60 genome specifies six related parameter values – one for each case – which we will refer to by adding case-numbers to the subscript (e.g.: $b_{min:1}$, $b_{min:2}$, ..., $b_{min:6}$). For ZIP60, the entire set of sixty parameters can be generated from the pattern $P_{t:n}$ where P is one of $\{m , b , g, c_a , c_r\}$; t is one of $\{min, D\}$; and n is an integer in $\{1,\ldots,6\}$. We'll refer to the set of six values for any one parameter-type (i.e., $\{P_{t:1}, P_{t:2}, \ldots, P_{t:6}\}$) as the *homologous set* of P_t parameter values. The ZIP60 interpretation for the six cases in each homologous set are as follows (where *q.type* indicates whether the last quote in the market was a bid or an offer):

- $P_{t:1}$: value of P_t for the case when a Seller raises its margin.
- $P_{t:2}$: value of P_t for the case when a Seller lowers its margin and *q.type=Bid.*
- $P_{t:3}$: value of P_t for the case when a Seller lowers its margin and *q.type=Offer*.

- $P_{t:4}$: value of P_t for the case when a Buyer raises its margin.
- $P_{t:5}$: value of P_t for the case when a Buyer lowers its margin and *q.type=Bid*.
- $P_{t:6}$: value of P_t for the case when a Buyer lowers its margin and *q.type=Offer*.

Finally, note that the additional computational costs of using ZIP60 as a replacement for ZIP8 are virtually zero. The space costs are those incurred in storing the additional 52 real-valued parameters: this is a large percentage increase, but in absolute terms it is still a very small amount of storage when expressed as actual additional bytes required. The additional time costs are also very low indeed: a tiny amount of extra processing is needed in initialising the ZIP60 trader (i.e., populating its look-up table of 60 real values) and then in doing table look-up while the trader is operating (i.e. choosing the values to use that are appropriate to the current "case"), but that's it. This is (almost) a free lunch.

3.2 ZIP60 results

In testing the performance of ZIP60 thus far, all effort has been devoted to exploring the performance of ZIP60 on the dual-shock tests: if markets populated by ZIP traders cannot cope with sudden shock-changes in supply and demand, then they are of very little practical interest. Moreover, it seems likely (but has not yet been empirically verified) that if ZIP60 does better than ZIP8 on these multi-shock tests, then it will also do better in those cases where there are fewer or no market shocks.

The experience now gained with GA optimisation of ZIP60 indicates that significant care is needed in managing the dimensionality of the search-space: simply applying the old methods that worked well with ZIP8 does *not* give best results when working with ZIP60. This is a lesson learnt from (partial) failure: for the very first attempts at evolutionary optimization of ZIP60 traders, the same experiment methods as described in Section 2 were used, except that the initial population was composed entirely of randomly generated ZIP60 individuals, rather than ZIP8s. The outcome of those experiments is illustrated in Figure 3, and comparison with Figure 2 reveals that the results from this first attempt with ZIP60 were somewhat mixed. Although ZIP60 EM scores were never worse than ZIP60 FM, and in some cases were significantly better, and while the scores of the elite evolved ZIP60 traders were *on the overall average* significantly better than the elite ZIP8 scores in the same experiments, the standard deviation on that average improvement was almost identical to the mean improvement itself. This very large standard deviation was a reflection of the fact that, in three specific cases (M434, M424, & M242), the evolved elite ZIP60 results were actually significantly *worse* than the corresponding ZIP8 results.

Now there is nothing preventing the ZIP60 GA system from evolving genotypes that correspond to ZIP8 solutions, so it seems peculiar that the ZIP60s perform worse than the ZIP8s in some cases. There are certainly points within the ZIP60 genome-space that correspond perfectly to ZIP8 solutions: if for each of the ten homologous sets the within-set variance of the parameter values for the set is (near) zero, then that ZIP60 genome is functionally equivalent to the corresponding single-case ZIP10 genome; and furthermore if the values of the $c_{a:D}$ and $c_{r:D}$ homologous sets are all zero, then the ZIP60 is functioning as a ZIP8. So, how come the ZIP60 results are sometimes worse than ZIP8? The fact that the GA failed even to find the known good ZIP8 solutions within the ZIP60 genome space is a strong indication that the 60-dimensional search space has characteristics (such as local maxima, sharp ridges, and plateaus in the fitness landscape) which make the search for good genomes a nontrivial process.

To address this, the ZIP genetic encoding was extended, allowing the number of cases (1, 2, 4, or 6, as discussed in Section 3.1) to be specified *on the genome itself*. The rest of the genome is still a set of ten homologous-set vectors (each made of six real numbers). If an individual’s gene specifying the number of cases is set to one, then all six parameter-values are set to be identical within each homologous set, by copying the values from the first element of the set into the remaining five. If the number of cases is set to two, then the three buyer-case parameter values within each set are forced to be identical copies of each other, as are the three seller values; and if the number of cases is set to be six, then the three buyer and the three seller parameters can all be different numeric values. Thus, the ZIP60 genomes are always 60 parameter-values long, but over-writing duplication of values within the genome can reduce the *effective* dimensionality of the parameter-vectors encoded on a particular genome so that it codes for *any* of the family of ZIP algorithms between ZIP60 and ZIP8. This is illustrated in Figure 4.

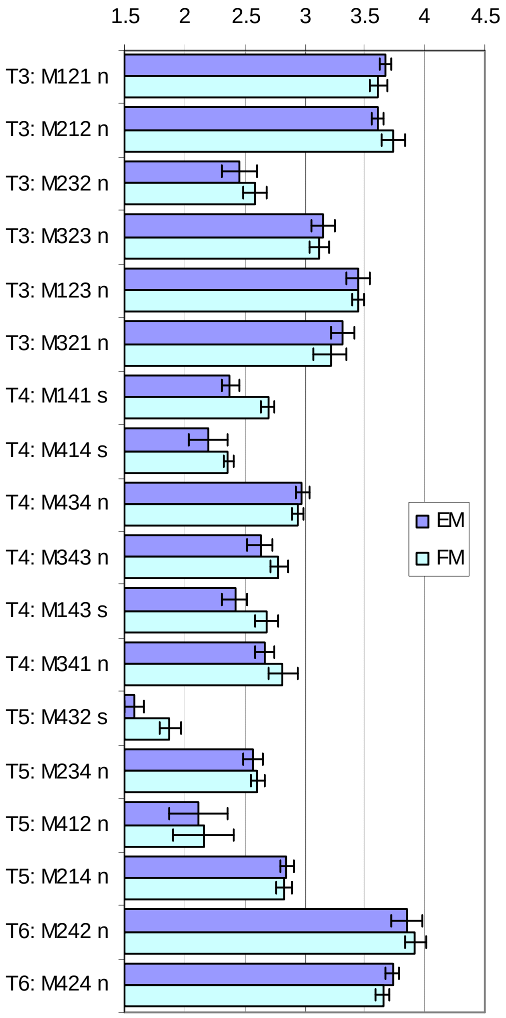

***Figure 3**: ZIP60 fitness data from the first attempt at GA optimization; format as for Figure 2.*

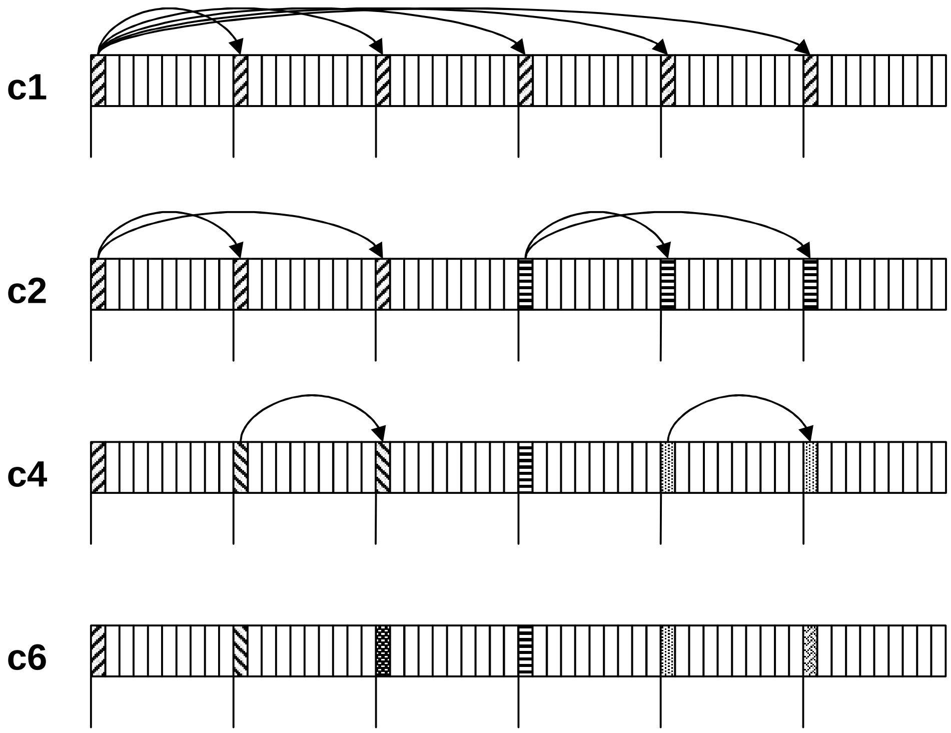

***Figure 4**: Schematic illustration showing how values are copied across the ZIP60 parameter vectors (i.e., genomes), which always have a length of 60, in the situations where the genetically-specified number of cases is (from top to bottom) 1, 2, 4, and 6 – labelled as c1, c2, c4, and c6, respectively. In each figure, the ZIP60 genome is shown as a line of 60 slots, with tick-marks marking the division into six groups of ten slots, one group for each possible case, so that homologous sets of parameters are distributed across the genome with successive values in each set being ten slots apart. In the c1 situation (top) the first set of ten values is systematically copied into each of the other five groups, so that each homologous set has the same value: this is indicated by the arrows showing the first value being copied into the six slots for that parameter's homologous set, and identical values are given the same shading-pattern. In the figures for c2, c4, and c6, the number of copy operations is reduced and so the number of (potentially) different values within one homologous set increases.*

The motivating hypothesis for placing the dimensionality of the search-space under evolutionary control was the belief that the GA's evolutionary search would be more successful if it could start by first simply optimizing the 1-case genome, and then (only once all the values are approximately correct) could successive multi-case refinements be progressively introduced by the GA as necessary. So, for example, if a 1-case individual mutated to become a high-case individual, thereby decoupling its genome-values across the different cases, such a mutant would only be retained in the population if the mutation that increases the number of cases is also associated with higher fitness. Strictly speaking, the initial case-increasing mutation is evolutionarily selectively neutral: the genome values for the different cases start out as identical copies of each other, but the case-increasing mutation allows *subsequent* mutations to introduce differences across cases, and it is those mutations that will be retained if they are correlated with higher fitness. Handing evolutionary control of the dimensionality of a search-space to the GA that is searching that space is an idea that was first explored in depth in Harvey's 1994 PhD thesis, where he provided detailed arguments for applying such *incremental evolution* and developed the "species adaptation genetic algorithm", which was first successfully applied in evolving neural-network controllers for autonomous mobile physical robots (Cliff, Harvey, & Husbands 1993).[9]

Two new sets of ZIP60 experiments were performed to test the effects of GA-controlled dimensionality. In the first set, the population was initialised with individuals that had a randomly-assigned value for the number of cases on their genome, with the values 1, 2, 4, and 6 being equally probable. This is the initialization we refer to here as ZIP60(1:6) (for "from 1 case to 6 cases"). In the second set, every individual in the initial population was set to have a 1-case genome; this is referred to here as the ZIP60(1:1) initialization. And so the first set of experiments, where all individuals in the initial populations were 6-case individuals, are referred to as ZIP60(6:6). Results from the EM experiments with ZIP60 with the (1:1), (1:6), and (6:6) initializations are shown in Figure 5, expressed as percentage changes over the corresponding mean ZIP8 EM scores from Figure 4.

Overall, on average, the ZIP60(1:1) scores are 14.0% better (lower) than the ZIP8 scores (and the s.d. on that mean improvement is 5.7%). In comparison, the ZIP60(6:6) scores are on average 12.91% better than the ZIP8, but the s.d. on that improvement is 12.88%; and for ZIP60(1:6), the average improvement is 12.32% with s.d.=7.03%. So, ZIP60(1:1) has the highest mean increase in performance and the lowest s.d. on its mean increase.

Results from significance analysis of the differences between the ZIP60(1:1) and ZIP60(1:6) upper-decile elite scores for the 18 dual-shock experiment schedules are tabulated in (Cliff, 2005), and they offer weakly supportive evidence for the claim that ZIP60(1:1) is a better initialization than ZIP60(1:6). Using the RRO test at the 1% significance level reveals that, over the 18 types of experiment, only for schedule M242 does ZIP60(1:1) lead to significantly better results than ZIP60(1:6). In all other cases, no statistically significant difference in the scores is detected. So, ZIP60(1:1) is certainly no worse than ZIP60(1:6), and the evidence thus far is that it is actually significantly better in one of the 18 cases studied. Moreover, the ZIP60(1:1) results are the only set of the three that improve on ZIP8 in *every* experiment. From this, we conclude that the (1:1) initialization method is to be preferred when using a simple GA to optimize a ZIP60 market. The absence of a huge difference between (1:1) and (1:6) is perhaps no surprise given that a ZIP60(1:1) system will, after sufficiently many generations, be pretty much indistinguishable from a ZIP60(1:6) as mutants with case-values greater than unity are progressively retained in the (1:1)-seeded population.

[9] A recent paper by Stanley & Miikkulainen (2004) re-discovers some of Harvey's (1994) ideas of evolutionarily controlled dimensionality increase and incremental evolution, which Stanley & Miikkulainen rename as "complexification".

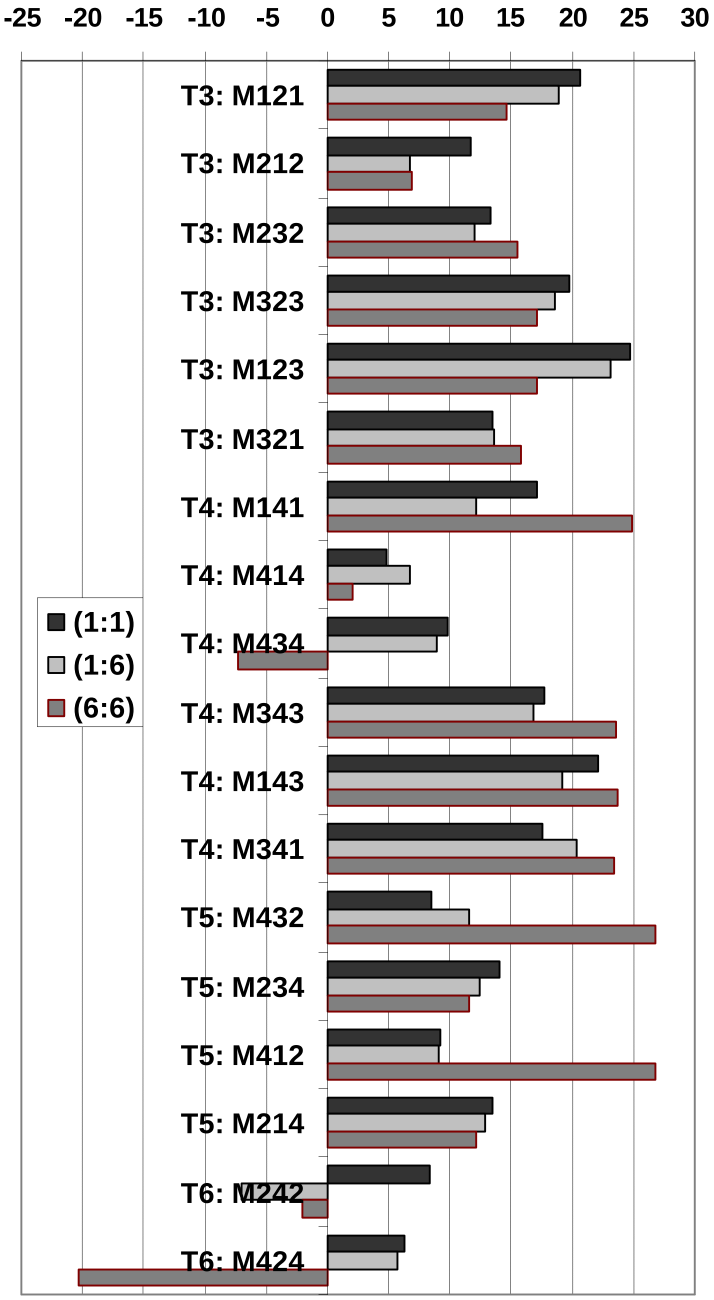

Figure 5: *Results from incremental evolution for the three different methods of initializing the ZIP60 population. Horizontal axis shows percentage improvement/deterioration, calculated as the difference between the ZIP60 and ZIP8 upper-decile mean elite scores, expressed as a percentage of the ZIP8 score. Positive values are improvements, negative values are deteriorations. See text for discussion.*

Figure 6 illustrates the underlying evolutionary dynamics in one set of 50 repetitions of a ZIP60(1:1) GA experiment. The four lines on the graph indicates, for each generation, the number of experiments in which that generation's elite individual's genotype is either 1-case, 2-case, 4-case, or 6-case (ZIP8, ZIP20, ZIP40, or ZIP60).[10] As can be seen, there is an initial rapid fall in the number of 1-case elite genotypes, as the number of elite genotypes that are 4-case and 6-case progressively rises. At approximately generation *g*=50, the number of 4-case elite genotypes levels off and then starts a steady decline which lasts until it has decayed to a noise-level at approximately *g*=300, but the number of 6-case genotypes continues to rise, and by g=300 pretty much all (49/50, i.e. 98%) of the experiments have an elite individual with a 6-case genotype.

The outcomes of the ZIP60(1:1) experiments that have been presented in this section are further analyzed and discussed in the next section.

[10] The source-code (Cliff, 2005) allows one-case genomes to be interpreted as either ZIP8 or ZIP10; in all the experiments reported in this paper, the interpretation was ZIP8.

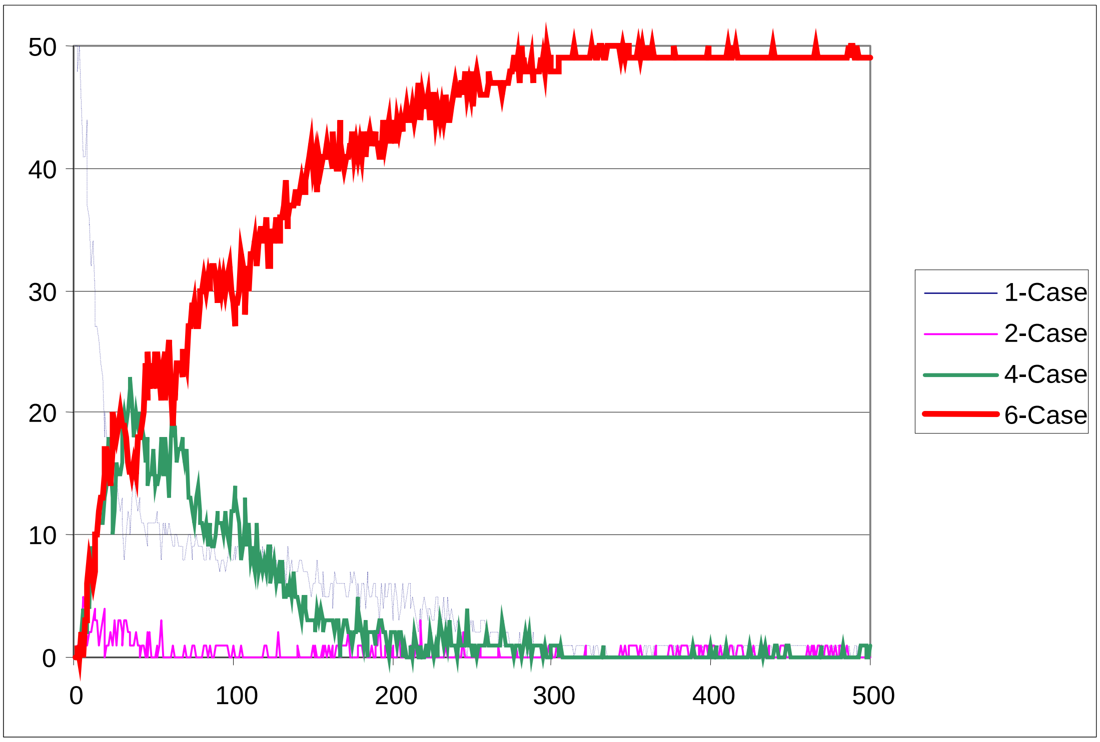

Figure 6: *Evolutionary dynamics of 50 repetitions (with different random seeds) of the M121 ZIP60(1:1) experiment. The four data series show, for each generation, the count (vertical axis) of the number of experiments in which the elite individual genotype at generation g (horizontal axis) has an allele value of 1, 2, 4, or 6 on its locus for the number of cases. Because this is a (1:1) initialization, the number of 1-case genotypes is initially 50, and for the three other types the number is zero. The number of 4-case and 6-case genotypes rises steadily for the first 50 generations, but thereafter the number of 4-cases goes into decline while the number of 6-cases continues increasing, to saturation at generation 300. The number of 2-case genotypes rises only very slightly in the early stages of the experiments (approximately the first 50 generations), and then falls back to noise levels.*

4. Analysis and Discussion

4.1 Analysis

The work reported in this paper is motivated by the belief that increasing the number of parameters in the ZIP algorithm improves the algorithm's performance. We need to be sure that no ZIP8 solutions are hidden in the evolved ZIP60 solutions.

That is, the results presented in Section 3 demonstrate that indeed ZIP60 out-performs ZIP8, and the fact that the GA typically settles on solution genomes with six cases (i.e., ZIP60) rather than four, two, or one (i.e. ZIP40, ZIP20, or ZIP8) indicates that a larger number of additional parameters are indeed useful; but (as was noted in Section 3.3) it is possible for a ZIP60 genome to be functionally equivalent to a lower-dimensioned ZIP*n* genome. In the most extreme case, if all the values in each homologous set of parameters are equal for any one genome (e.g. $b_{min1}=b_{min2}=b_{min3}=b_{min4}=b_{min5}=b_{min6}$), or if the differences between them are all sufficiently small to be ignored as mutational noise, then that ZIP60 genome is functionally equivalent to a ZIP10 genome (and if it has zero values for its c_{aD} and c_{rD} parameters, it is effectively a ZIP8). Of course, different homologous sets in the genome might have different values, but if for each homologous set the values within the set have (near-)zero variance then that ZIP60 genome is effectively a ZIP10. Different genomes may have different values at a particular locus, or in different homologous sets, but if all the genomes have the property that within each individual's genome the variance of the values in each homologous set is near-zero, then all those genomes are ZIP10.

So, to confirm that ZIP60 is indeed an advance on ZIP10 (or ZIP8), some analysis of the final evolved parameter-sets is necessary, to see whether they contain any low-dimensional solutions embedded in higher-dimensional spaces. To this end, principal component analysis (PCA)[11] was used on the parameter-values from the top-decile ZIP60(1:1) genomes (i.e. the genomes whose performance was summarised in Table 7). Specifically, each six-dimensional homologous set of final evolved parameter values was individually subjected to PCA, and the percentage of the variance in the parameter values accounted for by each principal component (PC) was calculated. If all the values in any one homologous set were equal or approximately equal, the first PC would account for close to 100% of the variance. However, the first PC would also account for close to 100% of the variance if the values in the homologous set were positioned along/around *any* line in the 6-D space, e.g. one where $P_{t:i}?P_{t:j}$ (for some P in $\{m , b , g, c_a , c_r\}$; some t in $\{min, D\}$; and for i, j integers in $\{1,\ldots,6\}$ with $i?j$). So, to identify a ZIP10 embedded in a ZIP60 genome, we'd need to see the first PC for each homologous set accounting for close to 100% of the variance, *and* see the angle q between the first PC and the line $P_{t:1}=P_{t:2}=\ldots=P_{t:6}$ being very close to zero. That is,

$$q = p - | \cos^{-1}((\mathbf{p.u}) / | \mathbf{p} |) - p | \qquad \text{(Equation 1)}$$

with $\mathbf{p}$ being the first PC (a 6-D vector) and $\mathbf{u}$ being a 6-D unit vector with elements $u_{i:i=1\ldots6}$ such that $u_1=u_2=\ldots=u_6>0$.

[11] PCA is explained in most textbooks on multivariate analysis: see, e.g., Chatfield & Collins (1980).

Figure 7 shows the results of the PCA analysis on the entire data-set of top-decile elite genomes, represented in a graphical format (the numeric data is tabulated in Appendix A). The first thing to note is that, although PC1 accounts for more than 50% of the variance in all homologous sets, the highest value is 90% for the PC1 of the b_{min} set, which is not high enough to cause any alarm. Additionally, the angle q is safely high in all cases: see Appendix A.

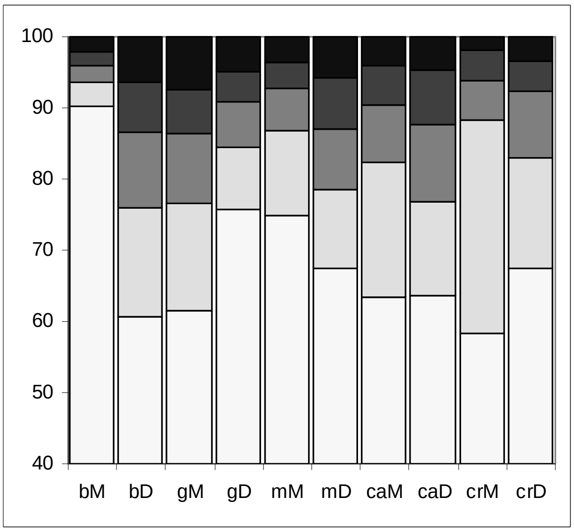

Figure 7: *Stacked-column histogram illustrating the PCA cumulative Percentage of Variance (PoV) explained by the six principal components PC1-PC6 for each homologous set of six variables, as defined in the text, for all 90 (=18*5) top-decile elite genomes from the ZIP60(1:1) experiments reported in Section 3; For each column, the lowest (and lightest-shaded) area represents the variance accounted for by the first PC; each successive darker-shaded area above that represents the variance accounted for by the next successive PC. The columns show, from left to right, the PCA data for* b_{min} *(labelled 'bM'),* b_D *('bD'),*g_{min} *('gM'),* g_D*('gD'),* m_{min} *('mM'),* m_D *('mD'),* $c_{a:min}$ *('caM'),* $c_{a:D}$ *('caD'),* $c_{r:min}$ *('crM'), and* $c_{r:D}$ *('crD').*

The second notable point in these data is that, for all homologous sets, PC6 makes no contribution to explaining the variance in the data, and typically the first four PCs account for over 95% of the variation. This indicates that the elite genomes occupy a 5-D subspace within the 6-D homologous set; and it's tempting to speculate that perhaps the only deviation from the 4-D subspace defined by PC1-PC4 in each case is mere noise. This is a point that will be returned to in the discussion of Section 4.2.

Although the data shown Figure 7 are encouraging, this is a rather coarse analysis because it considers all the ZIP60(1:1) elite genomes together, regardless of the treatment regime (i.e. what type of market shocks were applied). In Section 2.4, the four treatment regimes in (Cliff, 2003) were given the labels T3, T4, T5, and T6. The four histograms in Figure 8 illustrate the results of PCA on the values of the homologous sets of the top-decile elite genomes for each of these regimes treated separately (the numeric data is given in Appendix A).

The data shown in Figure 8 for T3, T4, and T5 experiments are all broadly similar to that in Figure 7: i.e., in all homologous sets the values of q are high and the first four PCs are needed to account for close to 100% of the variance. But the T6 data are very different: for T6, in the first six homologous sets (determining b_{min} , b_D , g_{min} , g_D , m_{min} , and m_D), the first PC alone accounts for over 98% of the variance, and the values of q (tabulated in Appendix A) are all very low: for these six homologous sets the T6 mean q is 1.05° (s.d.=0.52°) whereas for T3, T4, and T5 the mean q values are 16.32°, 17.15°, and 15.40° respectively (s.d.'s 6.37°, 7.00°, & 6.53°). This indicates that, for T6, the data-points in each of these homologous sets lie on lines in 6-D space very close to $\boldsymbol{u}$. Yet the PCs and q values for the last four T6 homologous sets (those affecting c_a and c_r) are much the same as those for T3, T5, and T5.

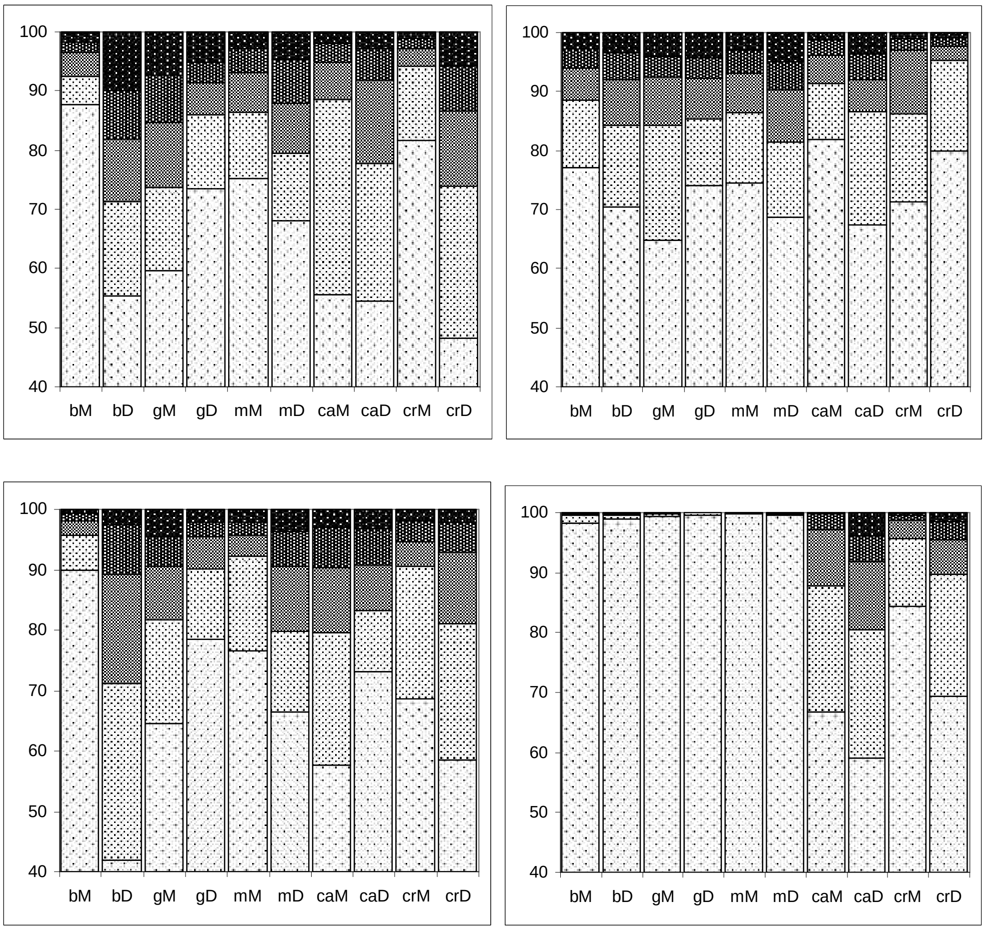

Figure 8: *Stacked-column histograms for the PCA PoV data from separate analysis of the T3, T4, T5, and T6 experiments (graphs at top left, top right, lower left, & lower right, respectively). Format as in Figure 7.*

To illustrate these points, Figure 9 shows the values of b_{min} from the final elite best-mode genomes evolved under the T6 treatments plotted as points in the 2D space defined by taking the first two principal components (PC1 & PC2) of that set as basis vectors. For comparison, we can project the same points onto a 2-D space defined relative to the vector **u** from Equation 1, so that each 6-D point **p** in the homologous set is projected to a 2-D point $(x_u\ ,\ y_u)$. To do this, let **u** denote the 6-D unit direction vector collinear with $[1,1,1,1,1,1]^T$. Then let x_u=**p.u** (i.e. the perpendicular projection of **p** onto **u**: that is, how far **p** comes along the line **u**) and let $y_u=|\mathbf{p} - x_u\mathbf{u}|$ (i.e. the perpendicular distance of the point **p** from the line **u)**. Projecting the 6-D b_{min} points onto this (x_u, y_u) space results in an arrangement of points that differs from Figure 9 only by two reflections: the data can be shown in an almost identical arrangement to that of Figure 9 by projecting each 6-D point to $(-x_u\ , -y_u)$, as shown in Figure 10.

Projecting each of the first six T6 homologous sets onto $(x_u\ , y_u)$-space gives a set of plots, all of which are qualitatively similar to Figure 10, as would be expected. This is a strong indication that, for the T6 experiments, we could potentially determine the 36 parameter-values in the first six homologous sets by specifying just one value for each set, and then for each set copying the value so that all parameters within the set hold the same value. So for the first six homologous sets in a ZIP60 genome tailored to T6 experiments, we'd only have to specify six independent values, and the remaining 30 parameter-values would be dependent on those six. And so we can reduce the (6*6=)36-D sub-space within ZIP60's 60-D space to a 6-D space of independent variables. But that still leaves the 24 values for the remaining four T6 homologous sets (i.e., those affecting c_a and c_r) to be determined. Intriguingly, it appears that there are additional regularities and dependencies in the elite evolved values for those sets that can be further exploited to reduce the number of independent variables that need to be specified.

As mentioned in Section 1, it is beyond the scope of this paper to conduct a full analysis and dimension-reduction of the evolved ZIP60 genotypes, but a sketchy proof-of-concept for the T6 genotypes can be further advanced here by examining Figures 11 and 12, which show 6-D homologous-set values for $c_{a:min}$ plotted in 2-D (x_u, y_u) space. Figure 11 shows the ten points from the T6 homologous set for $c_{a:min}$ which is analogous to the plot in Figure 10. It is clear in Figure 11 that the data points are not spread out on a flat line, but rather seem to be arranged along a diagonal line. Additional evidence for that interpretation comes in Figure 12, which shows the entire set of $c_{a:min}$ homologous sets (for all 18 market-shock regimes studied here) projected onto the same space: it is clear that there is a strong regularity in the elite evolved values for the $c_{a:min}$ parameter, and the best-fit straight line is very close to $y_u=x_u$ (with a high r^2). Qualitatively very similar plots result from similar projections onto (x_u, y_u)-space of the $c_{a:D}$, $c_{r:min}$, and $c_{r:D}$ elite evolved values as well. The regularities uncovered in these (x_u, y_u) plots, plus the PCA data, indicate that for each of the four homologous sets affecting c_a and c_r the evolved elite values lie on the lateral surface of a hypercone embedded in the 6-D space of that homologous set. Furthermore, the PCA data indicate that typically over 95% of the variance in the four c_a and c_r homologous sets can be explained by the first three principal components. So it's tempting to speculate that perhaps the good or optimal values of these parameters lies on the curved surface of a conventional 3D cone, appropriately positioned and oriented in the 6-space of each homologous set and with an appropriate vertex angle. If that is the case, it should be possible to derive a small set of independent variables from which the values for the 24 parameters in the c_a and c_r homologous sets can be determined. Further such analysis and dimensionality-reduction is beyond the scope of this paper, but can be addressed in future publications. The analysis presented here is sufficient to demonstrate that the pursuit of such lower-dimensional sets of independent variables is likely to be fruitful.

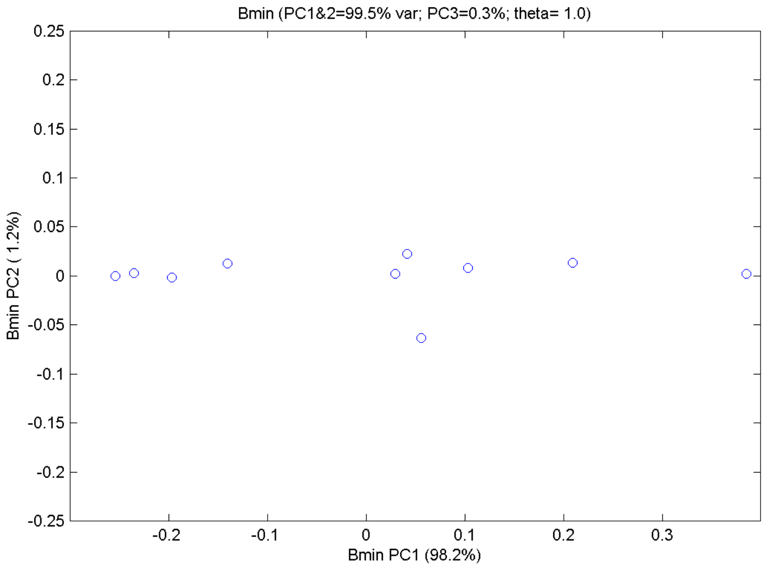

Figure 9: *Each 6-D point in the homologous set for the ten* b_{min} *values from the best elite-mode genotypes in the ZIP60(1:1) experiments with T6 market-shock regimes, projected onto the 2-D space defined by taking the first two principal components of that homologous set as basis vectors.*

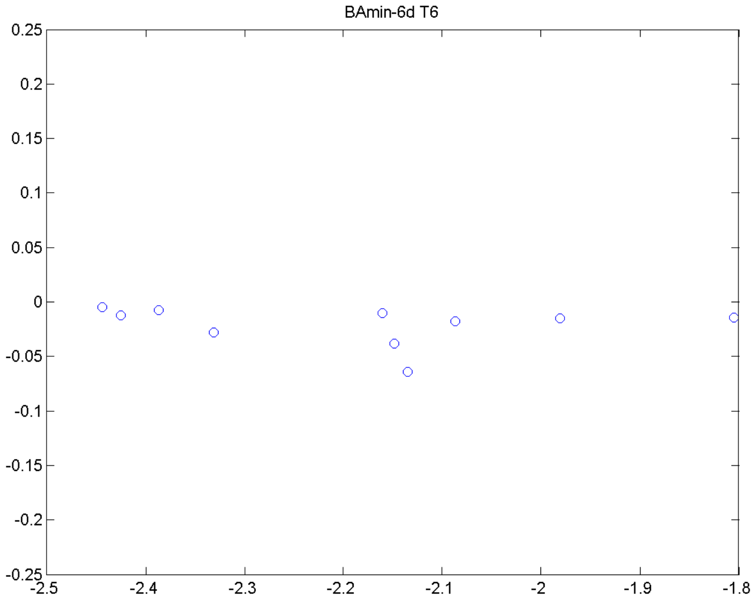

Figure 10: *The same set of 6-D points as in Figure 9, projected onto the 2-D* $(-x_u, -y_u)$ *space defined from the vector* **u** *in Equation 1; details of the projection are given in the text. Similarity with the projection in Figure 9 is obvious (and expected).*

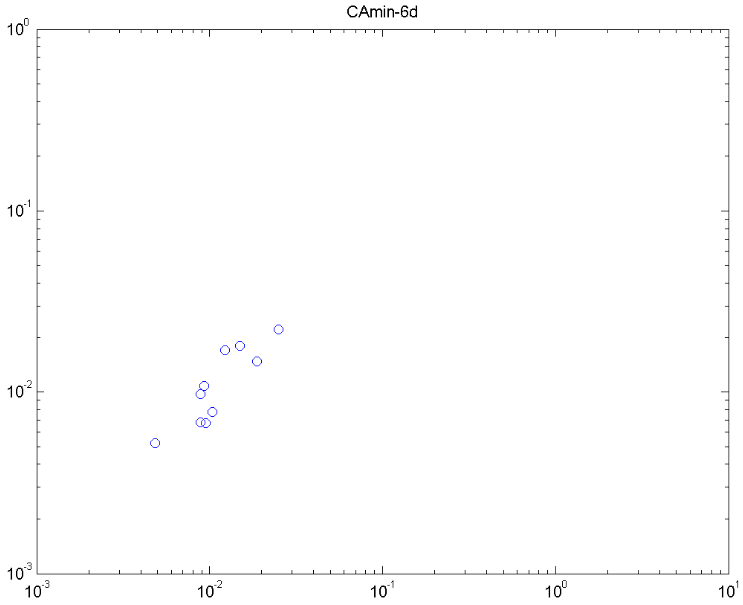

Figure 11: *Projection of the 6-D homologous set for the ten $c_{a:min}$ values from the top-decile best elite-mode genotypes in the ZIP60(1:1) T6 experiments onto the 2-D (x_u, y_u) space, as defined in the text. Although the axes in this plot are log-log, the data are approximately linear, and fit the line y_u=0.8552x_u+0.0014 with r^2=0.7647.*

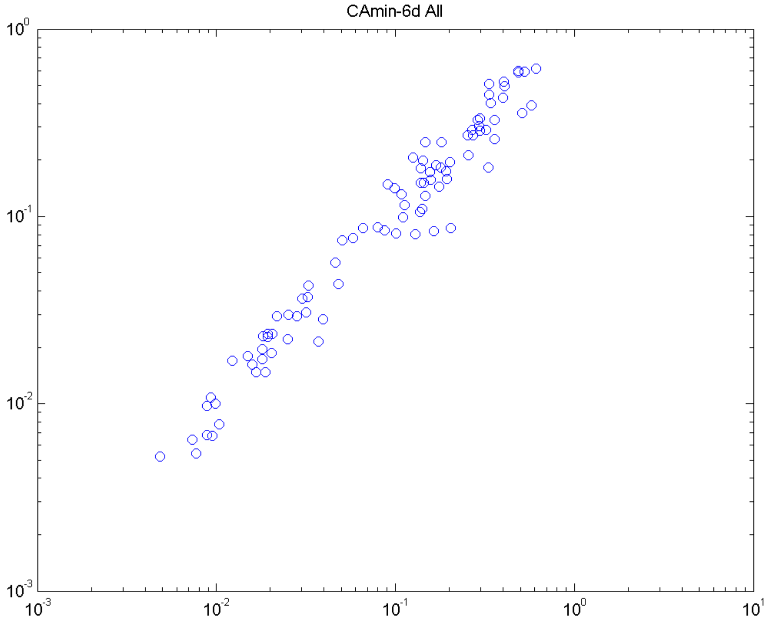

Figure 12: *Projection onto 2-D (x_u, y_u) space of the homologous set for the ninety $c_{a:min}$ values from the best elite-mode genotypes in the ZIP60(1:1) experiments from all 18 market-shock regimes. As with Figure 11, the data are approximately linear, and fit the line y_u=1.0097x_u+0.0035 with r^2=0.8926.*

4.2 Discussion and Further Work

Comparing the ZIP8 and ZIP60 results presented here reveals that with ZIP60 the GA much less frequently discovers hybrid values of Q_s yielding overall market dynamics that are better than those of the corresponding fixed-market CDA Q_s=0.5 experiments. That is, despite the final ZIP60 EM evolved Q_s values varying quite widely, few of them give results significantly better than the corresponding FM results. Data tables available in (Cliff, 2005) show that in two thirds (12 out of 18) of the original ZIP8 experiments, the EM experiment found a "hybrid" Q_s value that improved on the corresponding FM score; yet in comparable results from the ZIP60(6:6) experiments, the occurrence of superior EM results has fallen by 67%, i.e. from 12/18 down to 4/18 (similar falls occur with the (1:6) and (1:1) initialisations of the ZIP60 experiments). It seems likely that this is an indication that the previously-published results showing evolved hybrid auction mechanisms are to some extent artefacts of the lack of sophistication in the ZIP8 traders that were used in those studies. A counterpoint to this is that Byde (2003) presented results from applying similar GA-search for designs for hybrid sealed-bid auctions, where the GA found hybrid solutions to be preferable to the traditional first-price and second-price sealed bid auctions, and those results were independent of the sophistication of the traders in the market. So, the jury's out but there's some reasonable doubt, and this is an issue that should be explored in more depth in future research.

Figure 6 showed an illustrative evolutionary time-series for the frequency of each of the four possible case-types, over 50 repetitions of the same experiment. A notable feature in this graph is that the number of 2-case (ZIP20) genomes never rises very far. Although it is perhaps no surprise that the final elite genomes almost always encode for six-case ZIP60 solutions, the trajectories on this graph may indicate that the initial division into two cases, one for buyer-agents and one for seller-agents (a division made on the basis of intuition rather than any particular line of reasoning) confers no extra fitness of practical use – this extension of ZIP is pretty-much ignored by the GA, which jumps straight to the four-case and six-case extensions. This could be read as a clue that the narrative in Section 3.1, which is closely followed in the implementation of the genetic encoding (see the source-code in (Cliff, 2005)) may not be the best way of expanding from ZIP8 to ZIP16 (or from ZIP10 to ZIP20). The narrative in Section 3.1 could just as easily have been structured differently by making the *first* division into two cases be the one that distinguishes between situations in which margins are increased and situations in which margins are decreased; and then making the *second* division into four cases, thereby marking the additional distinction between buyers and sellers. That is to say, ZIP8 (and ZIP10) could just as easily be redefined to treat buyers and sellers equally, but to have different responses depending on in whether the margin is being adjusted up or down. In hindsight, a more sophisticated genetic encoding would allow each of the three case distinctions (buyer-*vs.*-seller; raise-*vs.*-lower; binary-*vs.*-ternary) to be varied independently. Working with that more sophisticated encoding might reveal new variants of ZIP with fewer than 60 parameters which nevertheless give results comparable to the ZIP60 results presented here.

Reducing the number of parameters without loss of performance with respect to ZIP60 is also possible in principle by use of multivariate analysis techniques, as was illustrated by the proof-of-concept dissection of the T6 elite genomes in Section 4.1. While this demonstrated the potential viability and usefulness of such dimensionality-reduction, it would be more attractive if an automated procedure could be developed. Such a procedure should take a collection of ZIP60 genomes and reduce the 60-D space to a smaller *v*-dimensional space of *independent* parameters, with the remaining (*60–v*)-dimensional set of *dependent* parameters having their values assigned automatically from appropriate functions of subsets of the independent parameters. There is a large and diverse literature covering techniques that could be brought to bear on this problem. One thing that the analysis in Section 4.1 makes clear is that simple linear techniques such as PCA are limited in their applicability, and so nonlinear alternatives to PCA are very likely to be needed: see (Hsieh, 2004) for a recent concise review of such nonlinear PCA approaches, and alternative techniques. In the same vein, mere analysis of correlations or variance between genome values and fitnesses within any one generation and also between successive generations is also something of a blunt tool: such evolutionary time-series probably do contain illuminating data, i.e. “signals” for interdependencies between parameters, but teasing those signals out from the mutational and behavioural “noise” may require more sophisticated approaches, such as analysis of cointegration and Granger causality, that are common in real-world financial data analysis (see, e.g., Alexander, 2001) but have not yet seen any serious use in studies of artificial trader agents.

Future papers can report on results from research into automatically setting and/or adjusting the sixty parameters in ZIP60 by exploiting dependencies between those parameters, and (eventually) on doing so without reliance on an offline/batch-mode optimizer such as the genetic algorithm used here. While operating in the 60-d space is no problem for an offline optimizer such as the GA, for online-real-time adaptation, operating in a space with fewer dimensions is likely to be much more efficient, because many algorithms for online and real-time adaptation and leaning scale poorly as the number of dimensions increases.

Finally, in all the experiments reported here, the GA used was the same simple (and inefficient) plain-vanilla GA first used in (Cliff, 1998) and also used for the experiments reported in all the other GA-ZIP papers (Cliff 2002a-d, 2003). The primary reason for using the same simple GA throughout this sequence of experiments was the desire to be able to easily compare results across the different styles of experiment reported on in that sequence of papers. But this strand of work has now matured to the point where the simplicity and inefficiency of the GA is a manifest weakness in the overall method. All that has been reported and analyzed in this paper is the final results of using the GA as a search and optimization process, and there are certainly several other automatic search and optimization techniques that could generate the same results quicker. Two grounds on which the GA implementation (Cliff, 2005) could be criticised are: the naïve way in which it deals with stochastic noise in the evaluation process; and the large amounts of information it throws away. The evaluation process deals with noise by simply taking an average over a large number of trials, but better methods have been known for a long time (e.g. Aizawa & Wah, 1994; Stroud, 2001). For simple GAs such as the one used here, a lot of information is thrown away when each generation is replaced by its successor (i.e. the immediate descendent population). The search is likely to be much more efficient if some type of Estimation of Distribution Algorithm (EDA) was used instead. In an

EDA, some or all of the genomes generated in the course of the GA are stored and analyzed, and the result of that analysis then influences or biases the subsequent generation of new genomes (Michalski, 2000; Larrañaga & Lozano, 2001; Zhang *et al.* 2004; Sun *et al.* 2005).

5. Conclusions

From the new data summarized and analyzed in this paper, it is clear that the ZIP60 variant of ZIP is a genuine improvement on the original ZIP8, and that ZIP60 parameter-vectors that outperform ZIP8 can be found by a simple GA so long as some care is taken in the application of that GA. Specifically, the care required is that the initial population be seeded with low-dimensional ZIP8s (or ZIP10s), and that expansion of the dimensionality of the search-space is allowed by the GA only when that expansion leads to increases in the fitness of the individuals concerned. Principal component analysis of the elite evolved parameter-sets from multiple runs under differently-changing supply and demand schedules revealed that the evolved parameter-vectors use considerably more values than the eight available in ZIP8, but also indicates that perhaps the best sets of parameter values reside in lower-dimensional subspaces within the 60-dimensional ZIP60 parameter space. Thus, it is possible that further analysis of elite evolved ZIP60 parameter-sets will reveal correlations and systematic dependencies, so that the full set of 60 parameter-values can be generated merely by specifying the values of an appreciably smaller number of independent parameters, from which the values of the remaining (dependent) parameters can then be set. Searching for such a subset of independent variables is likely to require nonlinear multivariate analysis techniques; and if such an independent sub-set can be identified then the next problem to be solved is finding a way to adjust those independent parameter-values "on the fly" as the market supply and demand alter dynamically, rather than relying on a "batch-mode" optimizer such as the GA used here. Finally, the fact that (in comparison to previous experiments using ZIP8 traders) the experiments with ZIP60 traders reported here show a reduced incidence of the discovery of "hybrid" auction mechanisms is possibly an indication that the hybrid auctions evolve as a consequence of a lack of sophistication in the behavior of ZIP8 traders: with the comparatively finer-grained responses of ZIP60 traders, hybrid mechanisms evolve much less frequently, and so it is tempting to conjecture that if the same type of auction-design experiments were repeated with even more sophisticated trader agents, hybrid mechanisms would not occur at all. Exploring that question remains one of several topics for further research.

Acknowledgements

Thanks to various colleagues and reviewers for their comments on previous incarnations of this paper. The data reported in Section 3 were generated while I was working at Hewlett-Packard Labs Bristol, which is where a very hurried first draft of this paper was written to accompany release of the source-code in (Cliff, 2005). Thanks to members of the HP Labs Complex Adaptive Systems research group for many valuable discussions over many happy years, especially to Andrew Byde and Martin Merry. The analysis in Section 4 was done while I worked as a Director on the Foreign Exchange Complex Risk desk of Deutsche Bank London, which is also where a more leisurely draft of this paper was subsequently refined. Thanks to my fellow FX Complex Risk quants and traders on the DB London FX trading floor for similarly valuable conversations and support, over several happy (but so very hectic) months, especially to Tony Hall and to Rhomaios Ram.

Appendix A: PCA Data Tables

Tables A.1 to A.5 tabulate the Principal Component Analysis (PCA) data that is summarized graphically in Figures 7 and 8, and they additionally show values for the angle q.

All	PC1	PC1&2	PC1-3	PC1-4	PC1-5	PC1-6	q
b_{min}	90.29	93.59	95.90	97.78	100.00	100.00	13.6
b_D	60.74	75.89	86.51	93.69	100.00	100.00	8.2
g_{min}	61.58	76.70	86.42	92.58	100.00	100.00	15.1
g_D	75.84	84.51	90.86	95.09	100.00	100.00	5.2
m_{min}	74.92	86.91	92.81	96.36	100.00	100.00	11.4
m_D	67.42	78.48	87.04	94.32	100.00	100.00	15.6
$c_{a:min}$	63.30	82.41	90.53	96.05	100.00	100.00	44.2
$c_{a:D}$	63.54	76.80	87.60	95.31	100.00	100.00	28.9
$c_{r:min}$	58.27	88.33	93.74	98.03	100.00	100.00	49.5
$c_{r:D}$	67.41	82.99	92.26	96.61	100.00	100.00	49.5

Table A1: *Cumulative percentage of variance (PoV) explained by the six principal components PC1-PC6 for each homologous set of six variables, as defined in the text, for all 90 (=18*5) top-decile elite genomes from the experiments reported in Section 3; right-most column shows the angle q (expressed in degrees), also defined in the text. This is the data illustrated in Figure 7.*

T3	PC1	PC1&2	PC1-3	PC1-4	PC1-5	PC1-6	q
b_{min}	87.63	92.54	96.63	98.25	100.00	100.00	15.9
b_D	55.31	71.25	81.89	90.04	100.00	100.00	11.4
g_{min}	59.55	73.63	84.62	92.73	100.00	100.00	15.4
g_D	73.42	86.04	91.29	94.85	100.00	100.00	8.4
m_{min}	75.27	86.49	93.04	97.18	100.00	100.00	20.7
m_D	67.99	79.60	87.99	95.19	100.00	100.00	26.1
$c_{a:min}$	55.54	88.47	94.72	98.03	100.00	100.00	53.7
$c_{a:D}$	54.50	77.85	91.81	97.14	100.00	100.00	25.4
$c_{r:min}$	81.58	94.09	97.18	98.85	100.00	100.00	57.1
$c_{r:D}$	48.28	73.85	86.67	94.27	100.00	100.00	33.8

Table A2: *Cumulative PoV explained by each PC, for the 30 top-decile elite genomes from the ZIP60(1:1) T3 experiments reported in Section 3 and illustrated in Figure 8.*

T4	PC1	PC1&2	PC1-3	PC1-4	PC1-5	PC1-6	q
b_{min}	77.07	88.57	93.89	97.25	100.00	100.00	23.7
b_D	70.44	84.16	91.93	96.48	100.00	100.00	11.8
g_{min}	64.88	84.18	92.49	95.87	100.00	100.00	27.0
g_D	74.12	85.31	92.28	95.60	100.00	100.00	9.8
m_{min}	74.62	86.46	93.18	96.98	100.00	100.00	18.1
m_D	68.68	81.49	90.26	95.00	100.00	100.00	12.5
$c_{a:min}$	81.88	91.40	96.01	98.74	100.00	100.00	53.2
$c_{a:D}$	67.44	86.58	92.08	96.34	100.00	100.00	42.6
$c_{r:min}$	71.26	86.17	97.06	99.02	100.00	100.00	54.8
$c_{r:D}$	79.89	95.18	97.58	99.17	100.00	100.00	56.0

Table A3: *Cumulative PoV explained by each PC, for the 30 top-decile elite genomes from the ZIP60(1:1) T4 experiments reported in Section 3 and illustrated in Figure 8.*

T5	PC1	PC1&2	PC1-3	PC1-4	PC1-5	PC1-6	q
b_{min}	89.98	95.73	98.00	99.28	100.00	100.00	16.1
b_D	41.88	71.18	89.28	97.50	100.00	100.00	24.1
g_{min}	64.60	81.76	90.59	95.46	100.00	100.00	14.5
g_D	78.45	90.13	95.54	97.77	100.00	100.00	7.7
m_{min}	76.46	92.29	95.66	97.78	100.00	100.00	8.8
m_D	66.41	79.84	90.46	96.24	100.00	100.00	21.2
$c_{a:min}$	57.63	79.47	90.23	96.95	100.00	100.00	32.7
$c_{a:D}$	73.21	83.17	90.83	96.84	100.00	100.00	32.3
$c_{r:min}$	68.67	90.45	94.66	98.13	100.00	100.00	53.0
$c_{r:D}$	58.59	81.17	92.95	97.91	100.00	100.00	42.0

Table A4: *Cumulative PoV explained by each PC, for the 20 top-decile elite genomes from the ZIP60(1:1) T5 experiments reported in Section 3 and illustrated in Figure 8.*

T6	PC1	PC1&2	PC1-3	PC1-4	PC1-5	PC1-6	q
b_{min}	98.23	99.47	99.74	99.95	100.00	100.00	1.0
b_D	98.91	99.62	99.81	99.91	100.00	100.00	0.8
g_{min}	99.36	99.83	99.93	99.97	100.00	100.00	2.1
g_D	99.60	99.94	99.97	99.99	100.00	100.00	0.7
m_{min}	99.88	99.96	99.99	100.00	100.00	100.00	0.8
m_D	99.61	99.86	99.93	99.98	100.00	100.00	0.9
$c_{a:min}$	66.69	87.86	97.26	99.68	100.00	100.00	52.7
$c_{a:D}$	59.00	80.49	91.80	96.22	100.00	100.00	50.0
$c_{r:min}$	84.43	95.71	98.72	99.60	100.00	100.00	64.3
$c_{r:D}$	69.31	89.78	95.59	98.44	100.00	100.00	64.9

Table A5: *Cumulative PoV explained by each PC, for the 10 top-decile elite genomes from the ZIP60(1:1) T6 experiments (where each market shock involves changes in both supply and demand) reported in Section 3 and illustrated in Figure 8.*

Appendix B: Details of the ZIP Algorithm

At the core of the ZIP algorithm is a minimally simplistic set of qualitative heuristics for adjusting trader *i*'s margin *m* depending on *i*'s current quote-price *p;* on whether *i* is still active in the market (i.e., still attempting to trade); and on whether the last quote in the marketplace *q* was an *offer* or a *bid*, and whether it was *accepted* or not. These are *heuristics* in the informal sense of the word: they are just some simple rules of thumb arrived at by introspection followed by some trial-and-error refinement, which have nevertheless been found to work extraordinarily well in practice. The heuristics for any one trader can be thought of as defining a simple decision tree with terminal nodes that specify whether the trader should raise or lower its margin, or default to leaving it unchanged. For the cases where a change in the margin is required, a set of quantitative steps for adjusting the trader's margin parameter is specified. Section B.1 shows pseudo-code for the heuristics; Section B.2 presents the equations that determine how the quantitative adjustments are made.

B.1 Qualitative Margin Heuristics

```
For SELLERS:
    – if ($q.accepted$ == true) then
        • forall sellers $s_i$
            – if $s_i.p = q.price$ then raise($s_i.m$)
        • if ($q.type$ == “bid”) then
            – forall active sellers $s_i$
                – if $s_i.p = q.price$
                – then lower($s_i.m$)
    – else
        • if ($q.type$ == offer) then
            – forall active sellers $s_i$
                – if $s_i.p = q.price$
                – then lower($s_i.m$)

For BUYERS:
    – if ($q.accepted$ == true) then
        • forall buyers $b_i$
            – if $b_i.p = q.price$ then raise($b_i.m$)
        • if ($q.type$ ==“offer”) then
            – forall active buyers $b_i$
                – if $b_i.p = q.price$
                – then lower($b_i.m$)
    – else
        • if ($q.type$ == bid) then
            – forall active sellers $b_i$
                – if $b_i.p = q.price$
                – then lower($b_i.m$)
```

B.2 Quantitative Margin Adjustments

The current quote-price at time t for trader i, denoted by $p_i(t)$, is set by a function of i's limit price l_i, and it's current margin $m_i(t)$:

$$p_i(t) = l_i \cdot (1+ m_i(t))$$

For sellers, $m_i(t) \in [0,8]\ \forall t$; to raise a seller's margin, the value of $m_i(t)$ is increased; and to lower the margin it is decreased. For buyers, $m_i(t) \in [-1,0]\ \forall t$; and raising a buyer's margin increases the (negative) magnitude of $m_i(t)$, taking it closer to -1.0, whereas lowering the margin reduces the magnitude, taking it toward zero.

ZIP uses the discrete-time Widrow-Hoff "delta" learning rule to adjust actual output A with respect to some desired output D, which is asymptotically approached at a rate determined by the "learning rate" parameter b:

$$A(t+1) = A(t) + D(t);$$
$$D(t) = b \cdot (D(t) - A(t)).$$

Which is modified to accommodate a momentum (damping) factor $g \in [0,1]$ as follows:

$$A(t+1) = g \cdot A(t) + ((1–g) \cdot D(t)).$$

For ZIP, the $A(.)$ updated by the Delta Rule is the profit margin m:

$$m_i(t+1) = (p_i(t) + \Delta_i(t)) / l_i – 1$$

And the D is defined relative to a desired "target price" t, as follows:

$$D_i (t) = b_i \cdot (t_I (t) – p_i(t))$$
$$t_i (t) = (\boldsymbol{A}_i(t) + \boldsymbol{R}_i(t) \cdot q(t))$$

Where $q(t)$ is the price of the last quote in the marketplace, and $\boldsymbol{A}(.)$ & $\boldsymbol{R}(.)$ are stochastic ***A***bsolute and ***R***elative perturbation functions, involving uniform distributions bounded by the constants c_a and c_r respectively, as were introduced in Section 2.1**.**

Combining all of the above thereby gives:

$$m_i (t +1) = (p_i(t) + G_i (t)) / l_i –1;$$

where $G_i(0)=0$,
and $G_i (t +1) = g_i \cdot G_i (t) + ((1–g_i) \cdot D_i(t)).$

For further details, see (Cliff, 1997).

Appendix C. The Genetic Algorithm

The simple GA used in (Cliff, 1998, 2001, 2002a-d, 2003) was again used here. Introductory texts on GAs include the good books by Goldberg (1989) and Mitchell (1998). In any one experiment, a population of size 30 was evolved for 500 generations: these GA parameter values were chosen on the basis of past experience. This choice of parameter values is uncontroversial, and the type of GA used here would be expected to give roughly the same results with different values of these two parameters, so long as the product of the two values is held roughly constant (i.e., so long as the total number of genomes evaluated is around 15000), and so long as the two values are roughly within an order of magnitude of each other. Each experiment was repeated 50 times with a different random seed on each repetition; it was commonly the case that the results from the 50 repetitions were multi-modally distributed.

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 real values for its ZIP parameters, plus an additional locus for the real-valued Q_s parameter, and latterly an additional locus (called *n_cases*) where the genome specifies the number of cases that it encodes. In each experiment, the initial random population was created by generating random real values from $\boldsymbol{U}[0,1]$ for each locus on each individual's genotype, except n_cases which is assigned an integer value of 1, 6, or randomly from {1,…,6} depending on the experiment. Crossover points were between the numeric values, and crossover was governed by a Poisson random process with an average of between one and two crosses per reproduction event. Mutation on the real-value loci 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 real 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: $log_{10}(m(g))=log_{10}(m_s)-((g/(999)).log_{10}(m_s/m_e))$, as was used in the GA of (Cliff, 1998, 2001, 2002a); where m_s is the "start" mutation limit (i.e., for $m(0)$); and m_e is the "end" mutation limit. In all the experiments reported here, as in all the previous GA-ZIP work, $m_s=0.05$ and $m_e=0.0005$. The mutation operator applied on *n_cases* was implemented by equi-probably choosing an integer from {1,…,6}.

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 (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 m_D, b_D, and g_D were clipped where necessary to satisfy the constraints $(m_{min}+m_D)<=1.0$, $(b_{min}+b_D)<=1.0$, and $(g_{min}+g_D)<=1.0$.

The fitness of genotypes was evaluated using the methods described previously in (Cliff, 1998, 2001, 2002a-d, 2003). 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 (often colloquially referred to as "days"), with allocations of stock and currency being replenished between each trading period. 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 days in the trial, using the method described in more detail in the body of this article. 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. So, because any one parameter vector specifies a set of distributions and is evaluated over 100 i.i.d. draws from that set of distributions, hence the evaluation is over the set of possible instantiations for ZIP60 traders defined by that set of distributions, rather than any one particular instantiation. Note also that as *minimal* deviation of transaction prices from the theoretical equilibrium price is desirable, so lower scores are better. That is, we aim here to *minimize* the evaluation scores: individuals with lower scores will on average have greater reproductive fitness (i.e., more offspring).

References

1. Aizawa, A. & Wah, B. (1994), “Scheduling of Genetic Algorithms in a Noisy Environment” *Evolutionary Computation*, **2**(2):97–122.

2. Alexander, C. (2001), *Market Models: A Guide to Financial Data Analysis.* John Wiley.

3. Bagnall, A., & Toft, I. (2004), “Zero Intelligence Plus and Gjerstad-Dickhaut Agents for Sealed Bid Auctions” presented at *Workshop on Trading Agent Design and Analysis (TADA04)*, New York, July 2004. (Preprint proceedings, pp.59–64), Available at http://www.eecs.harvard.edu/tada04/it-ajb-aamas04.pdf

4. Bagnall, A. & Toft, I. (2005) “Autonomous Adaptive Agents for Single-Seller Sealed Bid Auctions”, *Autonomous Agents and Multi Agent Systems.* In press.

5. Borgers, T. & Sarin, R. (1997), “Learning through reinforcement and replicator dynamics”, *Journal of Economic Theory*, **77**: 1–14.

6. Bowling, M. & Veloso, M. (2004), "Existence of Multiagent Equilibria with Limited Agents", *Journal of Artificial Intelligence Research* **22**:353-384,

7. 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.

8. Byde, A. (2003), “Applying Evolutionary Game Theory to Auction Mechanism Design”. *Proceedings of the 2003 ACM Conference on E-Commerce*. Also available as Hewlett-Packard Laboratories Technical Report HPL-2002-321.

9. Chatfield, C. & Collins, A. (1980), *An Introduction to Multivariate Analysis.* Chapman and Hall.

10. Clearwater, S., ed. (1995), *Market-Based Control*. World Scientific Press.

11. Cliff, D., Harvey, I., & Husbands, P. (1993), “Explorations in Evolutionary Robotics”, *Adaptive Behavior*, **2**:73–110.

12. Cliff, D. (1997), “Minimal-intelligence agents for bargaining behaviours in market environments”. Hewlett-Packard Laboratories Technical Report HPL-97-91.

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

14. Cliff, D. & Bruten, J. (1999), “Animat Market-Trading Interactions as Collective Social Adaptive Behavior". *Adaptive Behavior* **7**(3&4):385–414.

15. Cliff, D. (2001), “Evolutionary optimization of parameter sets for adaptive software-agent traders in continuous double-auction markets”. Presented at the *Artificial So-*

cieties 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.

16. Cliff, D. (2002a), "Evolution of market mechanism through a continuous space of auction-types". Presented at *Computational Intelligence in Financial Engineering (CIFEr)* session at *Congress on Evolutionary Computation (CEC2002)*, Hawaii, May 2002.

17. Cliff, D. (2002b), "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.

18. Cliff, D. (2002c), "Visualizing Search-Spaces for Evolved Hybrid Auction Mechanisms". Presented at the *Beyond Fitness: Visualizing Evolution* workshop at the 8th 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.

19. Cliff, D. (2002d), "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.

20. Cliff, D., Walia, V., & Byde, A. (2002), "Evolved Hybrid Auction Mechanisms in Non-ZIP Trader Marketplaces". *Proceedings International Conference on Computational Intelligence for Financial Engineering (CIFEr03)*, Hong Kong, March 2003.

21. Cliff, D. (2003), Explorations in evolutionary design of online auction market mechanisms. *Electronic Commerce Research and Applications* **2**(2):162–175, 2003.

22. Cliff, D. (2005), ZIP60: *Further Explorations in the Evolutionary Design of Online Auction Market Mechanisms.* Hewlett-Packard Laboratories Technical Report HPL-2005-85. Available from http://www.hpl.hp.com/techreports/2005/HPL-2005-85.pdf

23. 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).*

24. Easley, D., & Ledyard, J. (1992), "Theories of Price Formation and Exchange in Double Oral Auctions" in Friedman, D., & Rust, J. (editors) *The Double Auction Market: Institutions, Theories, and Evidence*, pp.63-97. Addison-Wesley;

25. Economist (2005), "The March of the Robo-Traders" *The Economist Technology Quarterly*, pp.23-24, 15 September 2005. Note: *The Economist* policy is to not identify the authors of their articles.

26. Feltovich, N. (2003), “Nonparametric tests of differences in medians: comparison of the Wilcoxon-Mann-Whitney and robust rank-order tests”, *Experimental Economics,* **6**:273–297.

27. Feltovich, N. (Forthcoming 2006), “Critical values for the robust rank-order test”, forthcoming, *Communications in Statistics.*

28. Friedman, D. (1991), “A Simple Testable Model of Price Formation in the Double Auction Market”, *Journal of Economic Behavior & Organization* **15**:47-70;.

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

30. Gerding, E., Somefun, K., & La Poutré, H. (2004), “Multi-Attribute Bilateral Bargaining in a One-to-Many Setting”, In *Proceedings of the Workshop on Agent Mediated Electronic Commerce VI (AMEC),* New York, USA.

31. Gjerstad, S. & Dickhaut, J. (1998), “Price Formation in Double Auctions”, *Games and Economic Behavior,* **22**:1–29.

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

33. Goldberg, D. (1989), *Genetic Algorithms: In Search, Optimization, and Machine Learning.* Reading, MA: Addison Wesley.

34. Gottlob, G., Greco, G., & Scarcello, F. (2005), "Pure Nash Equilibria: Hard and Easy Games", *Journal of Artificial Intelligence Research,* **24**:57-406.

35. Graham-Rowe, D. (2005), “How Bots Can Earn More Than You Do”. *New Scientist* (20 Aug., 2005) **187**(2513):26–27.

36. Greenwald, A., Guillemette, B., Naroditskiy, V., & Tschantz, M., (2005), “Scaling Up the Sample Average Approximation Method for Stochastic Optimization with Applications to Trading Agents” in Jansen, S. (ed) *Working Notes from the IJCAI-05 Workshop on Trading Agent Design and Analysis (IJCAI05:TADA05).* pp.21–27.

37. Harvey, I. (1994), *The Artificial Evolution of Adaptive Behavior*. PhD Thesis, School of Cognitive and Computing Sciences, University of Sussex, U.K.

38. He, M., Leung, H., & Jennings, N. (2003), "A fuzzy logic based bidding strategy for autonomous agents in continuous double auctions*" IEEE Transactions on Knowledge and Data Engineering* **15**(6):1345–1363.

39. Hsieh, W. (2004), Nonlinear Multivariate and Time Series Analysis by Neural Network Methods. *Review of Geophysics,* **42**(RG1003)1–25.

40. Jain, A., Murty, M. & Flynn, P. (1999), “Data Clustering: A Review”, *ACM Computing Surveys,* **31**(3)264–323.

41. Koza, J. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press.

42. Koza, J., Andre, D., Bennett, F., & Keane, M. (1998), *Genetic Programming: Volume 3*. Morgan Kauffman.

43. Ladley, D. & Bullock, S. (2005), “Who to listen to: Exploiting information quality in a ZIP-agent market” in Jansen, S. (ed) *Working Notes from the IJCAI-05 Workshop on Trading Agent Design and Analysis (IJCAI05:TADA05).* pp.28–34. Available from www.sics.se/tada05/tada05-proceedings.pdf.

44. Larrañaga, P. & Lozano, J. (eds) (2001), *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation.* Kluwer.

45. Li, L. & Smith, S. (2004), "Speculation Agents for Dynamic, Multi-period Continuous Double Auctions in B2B Exchanges", *Proceedings 37th Hawaii International Conference on System Sciences,* Hawaii, January 2004.

46. Lochner, K. & Wellman, M. (2004), “Rule-Based Specification of Auction Mechanisms.” *Proc. Third International Joint Conference on Autonomous Agents and Multiagent Systems.* pp.818-825.

47. Michalski, R. (2000), “Learnable Evolution Model: Evolutionary Processes Guided by Machine Learning”, *Machine Learning,* **38**(1-2):9–40.

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

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

50. Pardoe, D. & Stone, P. (2005), “Developing Adaptive Auction Mechanisms”, *ACM SIGecom Exchanges*, **5**(3):1–10.

51. Park, S., Durfee, E., & Birmingham, W. (2004), “Use of Markov Chains to Design an Agent Bidding Strategy for Continuous Double Auctions”, *Journal of Artificial Intelligence Research (JAIR),* **22**: 175–214.

52. Phelps, S., McBurney, P., Parsons, S., & Sklar, E. (2002), “Co-evolutionary auction mechanism design: a Preliminary Report”, in J. Padget, O. Shehory, D. Parkes, N. Sadeh, & W. Walsh (Eds.): *Agent-Mediated Electronic Commerce IV. Designing Mechanisms and Systems: AAMAS2002 Workshop on Agent-Mediated Electronic Commerce.* Lecture Notes in Computer Science Vol. 2531; Springer-Verlag.

53. Phelps, S., Parsons, S., & McBurney, P. (2004), *An Evolutionary Game-theoretic Comparision of Two Double Auction Market Designs.* Presented at the Fourth International Workshop on Agent-Mediated Electronic Commerce (AMEC-IV). Available from http://ana.lcs.mit.edu/peyman/amec-vi-accepted.htm

54. Phelps, S., Marcinkuiewicz, M., Parsons, S., & McBurney, P. (2005), “Using Population-Based Search and Evolutionary Game Theory to Acquire Better-response Strate-

gies for the Double Auction Market" in Jansen, S. (ed) *Working Notes from the IJCAI-05 Workshop on Trading Agent Design and Analysis (IJCAI05:TADA05).* pp.21–27. Available from http://www.sics.se/tada05/tada05-proceedings.pdf.

55. Preist, C. & van Tol, M. (1998), "Adaptive Agents in a Persistent-Shout Double Auction". *Proceedings of ICE-98,* ACM.

56. Pritchard, S. (2005), "Zippy Agents Going for Brokers". *Financial Times (London) FT-IT Review.* July 13, 2005: p.2.

57. Qin, Z. (2002), *Evolving Marketplace Designs by Artificial Agents.* MSc Thesis, Computer Science Department, Bristol University, September 2002.

58. Qin, Z. & Kovacs, T. (2004), "Evolution of realistic hybrid auctions" In: *Proceedings of the 2004 UK Workshop on Computational Intelligence (UKCI-04)*, pp.43–50. September 2004.

59. Reeves, D., Wellman, M., Mackie-Mason, J., & Osepayshvili, A. (2005), "Exploring Bidding Strategies for Market-Based Scheduling", *Decision Support Systems*, **39**:67–85.

60. Robinson, N. (2002), *Evolutionary Optimization of Market-Based Control Systems for Resource Allocation in Compute Farms.* MSc Thesis, School of Cognitive and Computing Sciences, University of Sussex, U.K.

61. Rothkopf, M., & Park, S. (2001), "An Elementary Introduction to Auctions", *Interfaces*, **31**:6(83–97).

62. Rumelhart, D., Hinton, G., & Williams, R. (1986), "Learning internal representations by error propagation", in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition*, Volume 1, edited by D. Rumelhart and J. McClelland (Cambridge, MA: MIT Press), pp. 318–362.

63. Rust, J., Miller, J., & Palmer, R. (1992), "Behavior of trading automata in a computerized double auction market" in Friedman, D., & Rust, J. (editors) *The Double Auction Market: Institutions, Theories, and Evidence,* pp.155-198. Addison-Wesley;

64. Shipp, D. (2004), *The effects of changes to supply and demand on trader agents and marketplaces.* MSc Thesis, School of Computing, University of Leeds, UK.

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

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

67. Stanley, K. & Miikkulainen, R. (2004), "Competitive Coevolution through Evolutionary Complexification", *Journal of Artificial Intelligence Research (JAIR),* **21**:63–100.

68. Stroud, P. (2001), “Kalman-Extended Genetic Algorithm for Search in Nonstationary Environments with Noisy Fitness Evaluations”. *IEEE Transactions on Evolutionary Computation*, **5**(1):66–77.

69. Sun, J., Zhang, Q., & Tsang, E., (2005), “DE/EDA: A New Evolutionary Algorithm for Global Optimization, *Information Sciences,* **169**(3):249–262.

70. Taylor, P., & Jonker, L., (1978), “Evolutionary Stable Strategies and Game Dynamics”, *Mathematical Biosciences*, **40**:145-156.

71. Tesauro, G. & Das, R. (2001), “High-Performance Bidding Agents for the Continuous Double Auction”, presented at *Economic Agents, Models, and Mechanisms Workshop, International Joint Conference on Artificial Intelligence (IJCAI),* August 2001.

72. Tesauro, G. & Bredin, J. (2002), “Strategic Sequential Bidding in Auctions using Dynamic Programming**”,** *Proceedings of the First International Conference on Autonomous Agents and Multiagent Systems (AAMAS-02)*, Bologna, Italy.

73. Vytelingum, P., Dash, R., David, E., & Jennings, N. (2004), "A risk-based bidding strategy for continuous double auctions" *Proceedings of the 16th European Conference on Artificial Intelligence* , Valencia , Spain, pp.79-83.

74. Walia, V. (2002), *Evolving Market Design,* MSc Thesis, School of Computer Science, University of Birmingham, UK.

75. Wellman, M., Reeves, D., Lochner, K., & Suri, R. (2005), “Searching for Walverine 2005” in Jansen, S. (ed) *Working Notes from the IJCAI-05 Workshop on Trading Agent Design and Analysis (IJCAI05:TADA05).* pp.1–6. Available from http://www.sics.se/tada05/tada05-proceedings.pdf.

76. Wichett, D. (2004), *Coadaptive Dynamics of Minimal Intelligence Trading Agents in Competing Populations.* MSc Thesis, Computer Science Dept, University of Birmingham, UK.

77. Wilson, R. (1987), "On Equilibria of Bid-Ask Markets" in Feiwel, G. (editor) *Arrow and the Ascent of Modern Economic Theory*, pp. 375-414. New York University Press.

78. Wilson, S. (1995), “Classifier Fitness Based on Accuracy” *Evolutionary Computation*, **3**(2):149–175.

79. Witt, U. (1992), Evolutionary concepts in economics. *Eastern Economic Journal,* **18**:405–419.

80. Wurman, P., Wellman, M., & Walsh, W. (2001), “A Parameterization of the Auction Design Space” *Games and Economic Behavior,* **35**:304-338.

81. Zhang, Q., Sun, J., Tsang, E., & Ford, J. (2004), “Hybrid Estimation of Distribution Algorithm for Global Optimization”, *Engineering Computations*, **21**(1):91–107.