Autonomous Agents and Multi-Agent Systems manuscript No.
(will be inserted by the editor)

What the 2007 TAC Market Design Game Tells Us About Effective Auction Mechanisms

(Long Version)

**Jinzhong Niu · Kai Cai · Simon Parsons ·
Peter McBurney · Enrico Gerding**

Received: February 9, 2009 / Revised version: October 5, 2009

Abstract This paper analyzes the entrants to the 2007 TAC Market Design Game. We present a classification of the entries to the competition, and use this classification to compare these entries. The paper also attempts to relate market dynamics to the auction rules adopted by these entries and their adaptive strategies via a set of post-tournament experiments. Based on this analysis, the paper speculates about the design of effective auction mechanisms, both in the setting of this competition and in the more general case.

Keywords Double auction · Mechanism design · Trading agent competition

1 Introduction

This paper is concerned with the Market Design game that was run as part of the Trading Agent Competition [44] (TAC) in July 2007. The Trading Agent Competitions have been held annually since 2000 with the aim of encouraging research into software agents that can bid for goods and services on behalf of their human owners [15, 20]. There have been several different games, but up until 2007 competing in these games had involved designing an agent that could bid effectively and make profitable transactions — the researchers who entered the games were, naturally, interested in how best to do this bidding. Our research, in contrast, is more concerned with the design of markets in which trading agents interact,

Jinzhong Niu · Kai Cai
Department of Computer Science, Graduate Center, City University of New York
365 Fifth Avenue, New York, NY 10016, USA
E-mail: {jniu, kcai}@gc.cuny.edu

Simon Parsons
Department of Computer and Information Science, Brooklyn College, City University of New York
2900 Bedford Avenue, Brooklyn, NY 11210, USA
E-mail: parsons@sci.brooklyn.cuny.edu

Peter McBurney
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK
E-mail: mcburney@liverpool.ac.uk

Enrico Gerding
Department of Electronic and Computer Science, University of Southampton, Southampton SO17 1BJ, UK
E-mail: eg@ecs.soton.ac.uk

and we introduced the Market Design game to encourage research in this area. The game was certainly successful in attracting entrants, and we had an exciting competition, but, as discussed below, there is not much that one can learn from the game itself.[1] The value of the competition is that it gives rise to a set of strategies that can be subsequently analyzed to extract general conclusions about how to approach problems like those in the competition. It is the aim of this paper to provide such an analysis.

1.1 Background

Auctions are special markets with restricted rules. Different auction designs may vary significantly in properties including efficiency, profit, and transaction volume. Well-designed auctions result in desired economic outcomes and are widely used in solving real-world resource allocation problems, and in structuring stock and futures exchanges. As a result, the field of auction mechanism design has drawn much attention in recently years from economists, mathematicians, and computer scientists [3, 19].

In traditional auction theory, auctions are viewed as games of incomplete information, and traditional analytic methods from game theory have been successfully applied to some *single-sided auctions*, where a single seller has goods for sale (or a single buyer desires to purchase goods) and multiple buyers bid for the goods (or sellers offer the goods), and some simple forms of *double auctions* (DAs), where there are multiple sellers and multiple buyers and both sides may make *offers* or *shouts*.

However, as, for example, Friedman [8] has pointed out, DAs, particularly *continuous double auctions* (CDAs),[2] are too complex to analyze in this way since at every moment, a trader must compute expected utility-maximizing shouts based on the history of shouts and transactions and the time remaining in the auction. This difficulty led researchers to seek experimental approaches. Smith [40] pioneered this field and showed, through a series of experiments with human subjects, that even CDAs with just a handful of traders can give high allocative efficiency and quick convergence to the theoretical equilibrium. Software agents armed with various learning algorithms and optimization techniques have been shown to produce outcomes similar to those obtained by human subjects [5, 14], are capable of generating higher individual profits [6], and can be used to explore the properties of auction mechanisms [53].

In parallel with the automation of traders, computer scientists have started to explore the automated design of auction mechanisms. Thus, Cliff [4] explored a continuous space of auction mechanisms by varying the probability of the next shout (at any point in time) being made by a seller, denoted by Q_s, and found that a Q_s that corresponds to a completely new kind of auction led to faster transaction price convergence. Phelps *et al.* [35] showed that genetic programming can be used to find an optimal point in a space of pricing policies, where the notion of optimality is based on allocative efficiency and trader market power. Niu *et al.* [25] presented a mechanism that minimizes variation in transaction price, confirming the mechanism through an evolutionary exploration. Pardoe and Stone [28] suggested a self-adapting auction mechanism that adjusts parameters in response to past results.

Although these evolutionary or adaptive approaches involve automatic processes, they make use of an array of candidate auction rules or parameterizable frameworks that are

[1] At the time of writing there have been two further Market Design games, in July 2008 and July 2009.

[2] A CDA is a continuous DA in which any trader can accept an offer and make a deal at any time during the auction period.

initially conceived by humans. Moreover, the result of an evolutionary exploration or an adaptive process, may depend on the quality of the candidate solutions which the process starts with — this was certainly the experience we had in [32] and [33].

When we started discussing the design of the Market Design game, our hope was to provoke further research in this form of mechanism design, concentrating on the continuous double auction. Previous studies usually present comparison of auction mechanisms in different proprietary settings which differ in terms of the information available to traders, computational resources and so on. As a result, mechanisms are difficult to compare, and we thought that offering a competition on a shared software platform would encourage the development of mechanisms that could be more easily compared. However, there was another aspect of existing work on double auctions that we wanted to address, that is the fact that all the work we were aware of considered single auctions, operating in isolation.[3] In contrast, not only do traders in an auction compete against each other, real markets face competition from other markets [39] and we wanted the Market Design game to reflect this kind of interaction. The format of the game we came up with is as follows. Each entrant in the competition provides a *specialist* that regulates a market with a set of auction rules, and these specialists compete against each other to attract traders and make profit. Traders in these games are provided by the competition platform and each of them learns to choose the best market to trade in. Because the Market Design game reverses the usual format of TAC competitions, we call it the CAT game.[4]

1.2 Strategy evaluation in competitive games

Trading competitions like CAT have been an effective tool in fostering innovative approaches and advocating enthusiasm and exchange among researchers [42,49]. However, the competitions themselves usually cannot provide a complete view of the relative strength and weakness of entries. In a competition, the performance of one player closely depends upon the composition of its opponents and the competition configuration, and the scenarios considered are usually limited. Thus we typically turn to post-competition analysis to tell us which entries are most interesting. Ideally, such an analysis will cover all possible scenarios, but this usually presents too large a possible space. As a result, a common practice is to deliberately select a limited number of representative strategies and run games corresponding to a set of discrete points or trajectories in the infinite space, assuming that the results are representative of what would happen in the whole space were one to explore it [41].

There are two common types of approaches to post-competition analysis: *white-box* approaches and *black-box* approaches. A white-box approach attempts to relate the internal logic and features of strategies to game outcomes. In the Santa Fe Double Auction Tournament and post-tournament experiments [37], a thorough examination of auction efficiency losses indicated that the success of the KAPLAN trading strategy is due to its patience in waiting to exploit other trading strategies. In Axelrod's Computer Prisoner's Dilemma Tournament [1], the strong showing of TIT FOR TAT is attributed to the fact that it is forgiving as well as being cooperative. While a white-box approach is often domain-dependent, the insights obtained in the concerned domain may still be extended to other domains. For instance, the payoff structure in the iterated Prisoner's Dilemma problem captures the nature of many other issues that are faced by parties with conflicting interests.

[3] Even work like [2,36] that compares two kinds of auction looks at the properties of each kind of auction operating in isolation.

[4] It is also the case that "catallactics" is the science of exchanges.

A black-box approach, on the other hand, considers strategies as atomic entities. One perspective is an *ecological* one based on *replicator dynamics*, from which the entities are biological individuals in an infinitely large population and a sub-population playing a particular strategy grows in proportion to how well that strategy performs relative to the whole population in average [11]. Walsh *et al.* [47] combines the game-theoretic solution concept of Nash equilibrium and replicator dynamics, turning a potentially very complex, multi-stage game of trading strategies into a one-shot game in normal form. What's more, a technique called *perturbation analysis* is used to evaluate whether a strategy can be improved further. Phelps *et al.* [31,32] successfully applied this approach in acquiring a better trading strategy for DA markets. Jordan *et al.* [16] took a similar approach to the evaluation of entries in the TAC Supply Chain Management Tournament (SCM) and other games [17].

1.3 Our contribution

This paper makes three main contributions. After a brief description of the game, it provides a classification of the entries based on their internal designs, and uses this classification to compare these entries. Since all the entries are double auction markets, this classification is a refinement of the classification presented in [52]. The paper then presents a white-box analysis of those entries to 2007 CAT competition (CAT 2007) that were available in the TAC agent repository,[5] and attempts to relate market dynamics to the auction rules adopted by these entries and their adaptive strategies through a set of post-tournament experiments. Finally, the paper performs a black-box analysis on the same set of specialists, examining the relative strength and weakness of the specialists in several scenarios, demonstrating some vulnerabilities in entries that placed highly in the competition.

This paper combines, revises and extends [22] and [23], in particular providing more explanation and additional results from the black-box analysis.

2 The Market Design game

2.1 Game procedure

A CAT game lasts a certain number of *day*s, each day consists of *round*s, and each round lasts a certain number of *tick*s, or milliseconds. Each game involves traders, which buy and sell goods, and specialists, which provide markets for those goods, enabling the trade. All traders and specialists are required to check in with the game server prior to the start of a game, and the list of all clients are broadcast to each client afterwards.

Before the opening of each day, the specialists are required to announce their price lists, which are then forwarded to all clients by the game server. After a day is opened, traders can register with one of the specialists (and only one specialist). Their choice of specialist depends on both the announced fees for that day, but also on the profits obtained in previous days. Traders will tend to choose specialists where they expect the highest profits. After a day closes, information on the profit by each specialist and the number of traders registered with it is disclosed, which allows specialists to adapt or learn to improve their competitiveness and eventually obtain higher scores.

Trading only takes place during a round. In a given round traders submit shouts to the specialists they are registered with and those specialists have the option to accept or reject

[5] `http://www.sics.se/tac/showagents.php`.

shouts. A shout that is accepted becomes active, and remains active until it is successfully matched with another shout or the trading day ends. A specialist may match *asks* (shouts to sell) and *bids* (shouts to buy) any time during a round, clearing the market. A matched bid must have a higher price than the corresponding ask, and the transaction price that is set must fall in between.

2.2 Traders

Each trading agent is assigned private values for the goods to be traded. The private values and the number of goods to buy or sell make the demand and supply of the markets. The private values remain constant during a day, but may change from day to day. Each trading agent is also endowed with a *trading strategy* and a *market selection strategy* to do two tasks respectively. One is to decide how to make offers, and the other is to choose market to make offers in. These two tasks allow our traders to exhibit intelligence in two, orthogonal, ways.

2.2.1 Trading strategies

Every trader uses one of the following four trading strategies, which have been extensively researched in the literature and some of them have shown to work well in practice:

- ZI-C (Zero Intelligence with Constraint): a simple strategy [14] which picks offers randomly but ensures the trader does not make a loss;
- RE (Roth and Erev): a strategy [7] that uses the profit earned through the previous shout as a reward signal and learns the best profit margin level to set, mimicking human game-playing behavior in extensive form games;
- ZIP (Zero Intelligence Plus): a strategy [5] that adapts its profit margin by using the Widrow-Hoff algorithm [50] to remain competitive in the market based upon information about shouts and transactions; and
- GD (Gjerstad and Dickhaut): a sophisticated strategy [13] that estimates the probability of an offer being accepted from the distribution of past offers, and chooses the offer which maximizes its expected utility.

ZIP and GD require information about the offers made by other traders and the results of those offers that ZI-C and RE do not need, and so traders that use these strategies may incur higher costs when specialists impose charges on shout and transaction information.

2.2.2 Market selection strategies

The market selection strategies that are possibly adopted by a trading agent include:

- random: the trader randomly picks a market;
- ε-greedy: the trader treats the choice of market as an n-armed bandit problem which it solves using an ε-greedy exploration policy [43]. An ε-greedy trader takes daily profits as rewards when updating its value function.
  An ε-greedy trader chooses what it estimates to be the best market with probability $1-\varepsilon$, and randomly chooses one of the remaining markets otherwise. ε may remain constant or be variable over time, depending upon the value the parameter α [43]. If α is 1, ε remains constant, while if α takes any value in $(0,1)$, ε will reduce over time.

– softmax: the trader is similar to an ε-greedy trader except that it uses a softmax exploration policy [43] in the n-armed bandit algorithm.
  Unlike an ε-greedy trader, a softmax trader does not treat all markets, other than the best market, exactly the same. If it does not choose the best market, it weights the choice of remaining market so that it is more likely to choose better markets. The parameter τ in the softmax strategy controls the relative importance of the weights a trader assigns markets, and similar to ε it may be fixed or variable controlled by α.

2.3 Specialists

Specialists facilitate trade by matching asks and bids and determining the trading price in an exchange market. Each specialist operates its own exchange market and may choose whatever auction rules for desired performance. Specialists are permitted and even encouraged to have adaptive strategies such that the policies change during the course of a game in response to market conditions. Section 3 presents a generic framework for discussing specialists in terms of the various policies that they implement.

A specialist can set its fees, or *price list*, which are charged to traders and other specialists who wish to use the services provided by the specialist. Each specialist is free to set the level of the charges (from zero up to some reasonable upper bounds). These are the following:

– *Registration fees*. Fees charged for registering with a specialist.
– *Information fees*. Fees for receiving market information from a specialist.
– *Shout fees*. Fees for successfully placing asks and bids.
– *Transaction fees.* A flat charge for each successful transaction.
– *Profit fees*. A share of the profit made by traders, where a trader's profit is calculated as the difference between the shout and transaction price.

The first four types of fees are each a flat charge, and the last one is a percentage charged on the profit made by a trader. A trader pays the registration and information fees at most once every trading day.

2.4 Assessment

The performance of specialists in a CAT game is assessed every day on multiple criteria. To encourage sustainable operation, not all the trading days will be used for assessment purposes, despite the fact that the game has a start-day and an end-day, and the selected assessment days are kept secret to entrants until they have been passed.

Each specialist is assessed on three criteria on each assessment day:

– *profit*: the profit score of a specialist on a particular day is given by the total profits obtained by that specialist on that day as a proportion of the total profits obtained by all specialists on that same day.
– *market share*: of those traders who have registered with a specialist on a particular day, the market share score of a specialist on that day is the proportion of traders that have registered with that specialist on that day.
– *transaction success rate*: the transaction success rate score for a specialist on a given day is the proportion of asks and bids placed with that specialist on that day which that specialist is able to match. In the case where no shouts are placed, the transaction success rate score is calculated as zero.

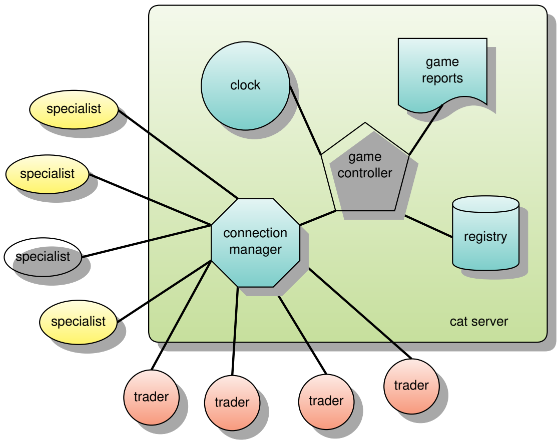


Fig. 1: The architecture of JCAT.

Each of these three criteria results in a value for each specialist for each day between 0 and 1. The three criteria are then weighted equally and added together to produce a combined score for each specialist for each assessment day. Scores are then summed across all assessment days to produce a final game score for each specialist. The specialist with the highest final game score will be declared the winner of the game.

2.5 Competition platform

JCAT [24], the platform that supports CAT games, extends the single-threading Java Auction Simulator API (JASA) [30], and adopts a client/server architecture. As Fig. 1 illustrates, the CAT server works as a communication hub, central time controller, and data logging facility, and CAT clients — either specialists or traders — communicate with each other via the server. On one hand, the CAT server takes traders' requests, including registering with a specialist, placing and modifying shouts, and forwards them to specialists; on the other hand, specialists notify the CAT server of matching shouts and, via the server, inform traders. The behaviors of the CAT server and CAT clients are regulated by the CAT Protocol, or CATP, which is detailed in [27]. The CAT server uses a registry component to record all game events and validate requests from traders and specialists. Various game report modules are available to process subsets of game events, calculate and output different metrics for post-game analysis.

3 Components of specialists

A specialist may adopt various auction rules. JCAT provides a reference implementation of a parameterizable specialist that can be easily configured and extended to use policies regulating different aspects of an auction. This section briefly describes a classification of those aspects that we have derived from the policies provided by JCAT and those used by specialists in the 2007 tournament. This classification is an extension of the parametric model of [52]. Section 4 relates these policies to the CAT 2007 finalists.

3.1 Matching policies

Matching policies define how a market matches shouts made by traders.

Equilibrium matching (ME) is the most commonly used matching policy [21,51]. The offers made by traders form the *reported demand and supply*, which is usually different from the *underlying demand and supply* that are determined by traders' private values and unknown to the specialist, since traders are assumed to be profit-seeking and make offers deviating from their private values. ME clears the market at the *reported* equilibrium price and matches intra-marginal asks with intra-marginal bids — with an intersecting demand and supply, the shouts on the left of the intersection (the equilibrium point) and their traders are called *intra-marginal* since they can be matched and make profit, while those on the right are called *extra-marginal*. Note that a shout, or a trader, that appears to be intra-marginal or extra-marginal in the reported demand and supply may not be so in the underlying demand and supply.

Max-volume matching (MV) aims to increase transaction volume based on the observation that a high intra-marginal bid can match with a lower extra-marginal ask, though with a profit loss for the buyer when compared with a match against an intra-marginal ask. A market using this form of matching is investigated in [9].

3.2 Quoting policies

Quoting policies determine market quotes issued by markets. Typical quotes are the ask quote and bid quotes, which respectively specify the upper bound for asks and the lower bound for bids that may be placed in a *quote-driven* market.

Two-side quoting[6] (QT) defines the ask quote as the minimum of the lowest tentatively matchable bid and lowest unmatchable ask, and defines the bid quote as the maximum of the highest tentatively matchable ask and highest unmatchable bid.

One-side quoting (QO) is similar to QT, but considers only the standing shouts closest to the reported equilibrium price from the unmatched side. When the market is cleared continuously (see below), QO is identical to QT.

3.3 Shout accepting policies

Shout accepting policies determine if a shout made by a trader should be entered in the market.

Always accepting (AA) accepts any shout.

Quote-beating accepting (AQ) allows only those shouts that are more competitive than the corresponding market quote. This is commonly used in both experimental settings and real stock markets, and is sometimes called "New York Stock Exchange (NYSE) rule" since that market adopts it. Clearly there is an interaction between such a policy and the quoting policy used by the market.

[6] The name follows [21] since either quote depends on information on both the ask side and the bid side.

Equilibrium-beating accepting (AE) estimates the equilibrium price based on past transaction prices, and requires bids to be higher than the estimate and asks to be lower. This policy was suggested in [25] and found to be effective in reducing transaction price fluctuation and increasing allocative efficiency in markets populated with ZI-C traders [14].

Self-beating accepting (AS) accepts all first-time shouts but only allows a trader to modify its standing shout with a more competitive price. AS imposes a looser restriction than AQ for extra-marginal shouts, but a tighter one for intra-marginal shouts since traders have to beat their standing shouts which are already more competitive than the corresponding market quote.

Transaction-based accepting (AT) tracks the most recently matched asks and bids, and uses the lowest matched bid and the highest matched ask to restrict the shouts to be accepted. In a *clearing house auction* (CH)[7] [10], the two bounds are expected to be close to the estimate of equilibrium price in AE, while in a CDA, AT may produce much looser restriction since extra-marginal shouts may steal a deal.

History-based accepting (AH) is inspired by the GD trading strategy. GD calculates how likely a shout is to be matched to determine what shouts to make. AH makes the same calculation and only accepts shouts that will be matched with probability no lower than a specified threshold. It is named after its need for the history of shouts and transactions in the market. Appendix A describes AH in detail as part of a simple, but powerful, market mechanism for competing in CAT games.

3.4 Clearing conditions

Clearing conditions define when the market is cleared and transactions are executed.

Continuous clearing (CC) attempts to clear the market whenever a new shout is placed.

Round clearing (CR) clears the market after all traders have submitted their shouts. This was the original clearing policy in NYSE, but was replaced, in the mid 1860s, by CC in order to generate immediate transactions and handle increased volumes. With CC, an extra-marginal trader may have more chance to steal a deal and get matched.

Probabilistic clearing (CP) clears the market with a predefined probability, p, whenever a shout is placed. It thus defines a continuum of clearing rules with CR ($p = 0$) and CC ($p = 1$) being the two ends.

3.5 Pricing policies

A pricing policy is responsible for determining transaction prices for matched ask-bid pairs. The decision making may involve only the prices of the matched ask and bid, or more information including market quotes.

Discriminatory k-pricing (PD) sets the transaction price of a matched ask-bid pair at some point in the interval between their prices. The parameter $k \in [0,1]$ controls which point is used and usually takes value 0.5 to avoid a bias in favor of buyers or sellers.

Uniform k-pricing (PU) is similar to PD, but sets the transaction prices for all matched ask-bid pairs at the same point between the ask quote and the bid quote. PU cannot be used

[7] A CH is another common type of DA. Unlike the CDA it clears at a pre-specified time, allowing all traders to place offers before any matches are found. A CH is used, for example, to set stock prices at the beginning and the end of trading on some stock exchanges [38].

with MV because the price intervals of some matched ask-bid pairs do not cover the spread between the ask quote and the bid quote.

n-pricing (PN) sets the transaction price at the average of the latest *n* pairs of matched asks and bids. If the average falls out of the price interval between the ask and bid to be matched, the nearest end of the interval is used. This policy, introduced in [25], can help reduce transaction price fluctuation and has little impact on allocative efficiency.

Side-biased pricing (PB) is basically PD with *k* set to split the profit in favor of the side on which fewer shouts exist. Thus the more that asks outnumber bids in the current market, the closer *k* is set to 0.

3.6 Charging policies

Charging policies determine how charges are imposed by a specialist. Specific strategies provided in the JCAT source code (and explored in [26]) are the following.

Fixed charging (GF) sets charges at a specified fixed level.

Bait-and-switch charging (GB) makes a specialist cut its charges until it captures a certain market share, and then slowly increases charges to increase profit. It will adjust its charges downward again if its market share drops below a certain level.

Charge-cutting charging (GC) sets the charges by scaling down the lowest charges of markets imposed on the previous day. This is based on the observation that traders all prefer markets with lower charges.

Learn-or-lure-fast charging (GL), adapts its charges towards some desired target following the scheme used by the ZIP trading strategy. If the specialist using this policy believes that the traders are still exploring among specialists and have yet to find a good one to trade, the specialist would adapt charges towards 0 to lure traders to join and stay; otherwise it learns from the charges of the most profitable market. GL uses an exploring monitor component to determine whether traders are exploring or not. A simple exploring monitor, for example, examines the daily distribution of market shares of specialists. If the distribution is flat, the traders are considered exploring, and not otherwise. This is based on the observation that traders all tend to go to the best market and cause an imbalanced distribution. Another scheme for the exploring monitor is to check the trader distribution in the latest several days and uses the relative market share gain and loss to determine whether it is good to lure traders.

3.7 Traditional double auction mechanisms

The policies presented in the previous section can be combined to easily create auction mechanisms, including those commonly used. Without considering the charging component, a CDA can be represented as

$$\mathsf{ME}+\mathsf{QT}+\mathsf{AQ}+\mathsf{CC}+\mathsf{PD} \tag{1}$$

while a CH can be represented as

$$\mathsf{ME}+\mathsf{AA}+\mathsf{CR}+\mathsf{PU} \tag{2}$$

4 Characteristics of specialists in the first TAC CAT Competition

The first CAT competition was held in conjunction with AAAI in July 2007. Table 1 lists the finalists in descending order of their final rankings[8] and identifies the auction rules we inferred from the programs of the CAT 2007 competition final (held in the TAC repository) against the policies we described in Section 3. All specialists for which we have data fit into the generic double auction mechanism framework introduced above and Table 1. [9]

We found that most specialists in the competition used ME to clear markets at the equilibrium price. `IAMwildCAT` and `Mertacor` were the only two attempting to match competitive intra-marginal shouts with extra-marginal shouts close to the equilibrium point in order to obtain high transaction success rates. QT, familiar from classic CDAs and CHs, is a popular quote policy, but its effectiveness is bound to the matching policy that is used with it since different matching algorithms, such as ME and MV, can generate significantly varying quotes. Furthermore, quote policies only affect the performance of the specialists when AQ is used as an accepting policy.

Specialists use a wide range of shout accepting policies, which reflects the importance of this aspect in performing well in CAT games. In contrast, only `CrocodileAgent` and `Mertacor` use a clearing condition that isn't one of the standard policies provided in JCAT.

Since JCAT ensures that specialists impose uniform charges on all traders registered with it on a trading day, it is not possible to attract specific traders by levying differential charges. However, about half the entrants managed to bias their pricing policy to promote the quality of their trader population.

Entrants seem to have contributed more effort to charging policies than to any other aspect of auction mechanisms. Table 3 in particular compares:

1. How charges are updated over time.
   Some specialists *adapt* their charges while others *directly calculate* the charges that they expect to bring a certain payoff without explicitly considering how they charge currently. A third choice is to combine the two approaches by setting charges that move *gradually* from the current level to the target level.
2. Whether different types of charges are treated differently.
   About half of the specialists impose only or mainly registration fees and charges on profits. `TacTex` charges only shout fees. `CrocodileAgent`, `Havana` and `MANX`, which don't have a bias towards a particular kind of fee, adapt charges without using any heuristic knowledge of the fee types.
3. Whether traders are identified and treated differentially.
   Only `IAMwildCAT` tracks individual traders and records information on them.
4. How much profit a trader and/or a specialist can make on average.
   `IAMwildCAT` and `jackaroo` are the only two specialists that lay down a road map for achieving some desired or target profit. `IAMwildCAT` is the only one that tracks the absolute value of the daily overall profit of specialists, which, when small, can be exploited by the specialist to obtain a fairly high share of the profit without imposing massive fees.
5. Whether a specialist learns from the history of charges and performances of its own and/or the other specialists.

[8] Due to technical problems, two teams, `TacTex` and `MANX`, were not able to participate in all the games. Some teams were banned from parts of some games — `PSUCAT` and `Havana` for exceeding reconnection limits, and `CrocodileAgent`, `Havana`, `MANX`, `PSUCAT`, `TacTex`, and `jackaroo` for invalid fees.

[9] Subsequent to the analysis undertaken here, two teams have reported on their specialists [29, 46].

Table 1: Comparison between the CAT 2007 finalists.

market	matching	quoting	accepting	clearing	pricing	charging
`IAMwildCAT`	ME+MV*	QT + QO + Δ	AQ+AE+AS+Δ	CR	PB*	Δ
`PSUCAT`	ME	(QT)	AE*	CC	PB*	Δ
`CrocodileAgent`	ME	(QT + QO*)	AE	CR*	PN*+PB*	GL*
`jackaroo`	ME	QT*	AQ	CC	PN	GC*+Δ
`Havana`	ME*	QT	AQ	CC	PD	Δ
`PersianCat`	ME*	(QT)	AT*+Δ	CC	PD	GF*+Δ
`Mertacor`	MV*	(QT)	AE*	CR*	PB*	Δ
`TacTex`	ME	(QT)	AA	CR	PD	GB*+GC*
`MANX`	ME	QT	AQ	CR	PD	GC*+GL*

XX^* denotes a policy that can be viewed as a modified or improved XX; Δ stands for some mechanism that cannot be related to any policy in Section 3; (XX) represents a quote policy that is defined by the specialist but has no effect on its behavior due to its adoption of a non-AQ accepting policy; and $XX + YY$ means some combination of XX and YY. `PhantAgent` is not included since it is not in the TAC repository.

Table 2: The scores of specialists in our experiments. The order follows the ranking in the 2007 competition.

specialist	**score**	**std. dev.**
IAMwildCAT	240.22	2.82
PSUCAT	209.26	12.01
CrocodileAgent	179.64	17.53
jackaroo	182.80	24.30
PersianCat	128.82	5.57
Mertacor	100.11	8.57
TacTex	166.66	8.99
MANX	140.09	31.03

It is a common practice among the specialists for fees to be set based on information about their competitors' charges and performances, though the lengths of history monitored vary from only the previous day, to a sliding multi-day window, to the full game history.

6. Whether a specialist tries to lure traders by charging less in the early stage of a game (*start effect*) and/or imposes higher charges when the game is about to end (*deadline effect*).
   Most specialists feature start and deadline effects, taking advantage of a definitive game duration and traders exploring widely at the beginning of a CAT game.

The characterization in Table 3 is a first step in establishing relationship between auction rules and auction performance. The next step is to start to identify the effects of these rules.

5 A white-box analysis of CAT 2007 entries

To further examine the specialists that participated in the CAT 2007 competition, we ran a series of games with the same setup as in the 2007 final games.

5.1 Experimental setup

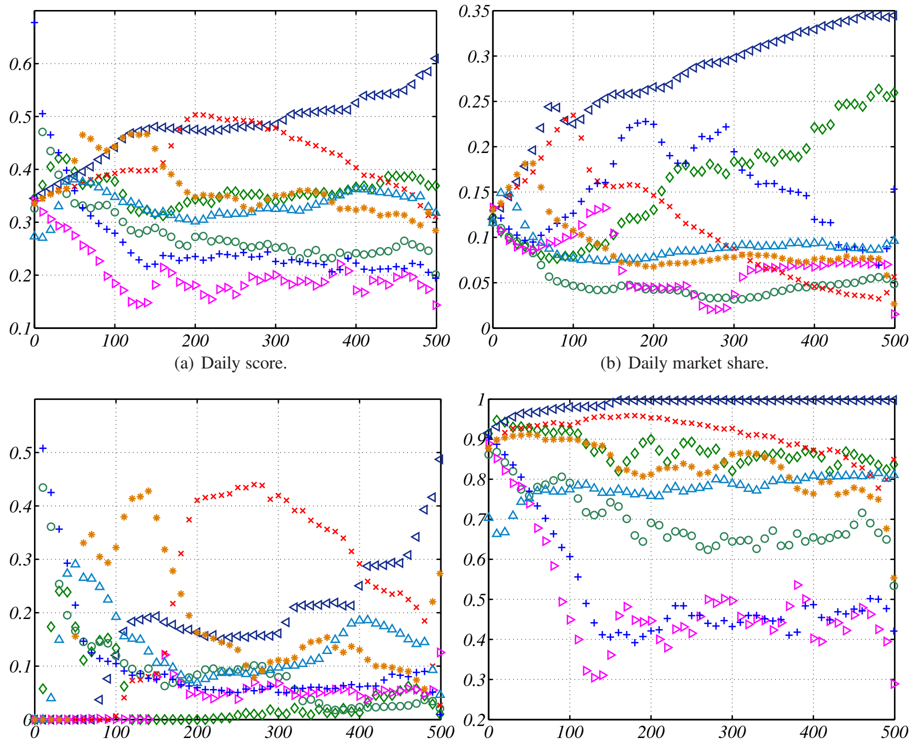
Every game in our experiment ran for 500 trading days with 10 1-second rounds per day. There were 180 ZIP traders, 180 RE traders, 20 ZI-C traders, and 20 GD traders. For each type of trader there were an equal number of buyers and sellers. The private values of all the traders were independently drawn from a uniform distribution between 50 and 150, and each trader was allowed to buy or sell up to three commodities per day. The specialists in our games include all eight of the 2007 specialists in the repository on the TAC website that we were able to run — Havana, which is in the repository, requires the CPLEX library which we do not have access to. The game server and all the clients were run on a single machine, a different setup from the CAT 2007 final games where entrants ran their specialists on machines that connected to the game server over the Internet. We used the same scoring criteria as in the tournament [12] (these were briefly described in Section 2.4), but, unlike the tournament, all the game days were assessed. The results and plots shown in the following

[10] PSUCAT however does identify traders to adjust parameters in its pricing policy.

Table 3: Comparison between the charging policies of the CAT 2007 finalists.

market	fee update	fee type bias	trader id	profitability		fee history		score history		start effect	deadline effect
				traders	specialists	self	others	self	others		
`IAMwildCAT`	⇒o	✓	✓	✓	✓	▮	▮	◇	◇	✓	✓
`PSUCAT`	o⇒	✓	✗[10]	✗	✗	▮	▮	▮	▮	✓	✓
`CrocodileAgent`	o⇒o	✗	✗	✗	✗	◇	◇	◇	◇	✓	✗
`jackaroo`	⇒o	✓	✗	✓	✗	▮	▮	▮	▮	✓	✓
`Havana`	⇒o	✗	✗	✗	✗	◇	◇	◇	◇	✓	✗
`PersianCat`	o⇒	✓	✗	✗	✗	▮	✗	▮	✗	✓	✓
`Mertacor`	⇒o	✓	✗	✗	✗	▮	◀	▮	◀	✓	✓
`TacTex`	o⇒	✓	✗	✗	✗	▮	▮	◀	◀	✓	✗
`MANX`	⇒o	✗	✗	✗	✗	▮	▮	◇	◇	✗	✗

✓	has this feature	✗	does not have this feature		
◇	sliding window	▮	single day	◀	full history
o⇒	adapting	⇒o	direct calculation	o⇒o	gradual learning



(a) Daily score. (b) Daily market share.

(c) Daily profit share. (d) Daily Transaction success rate.

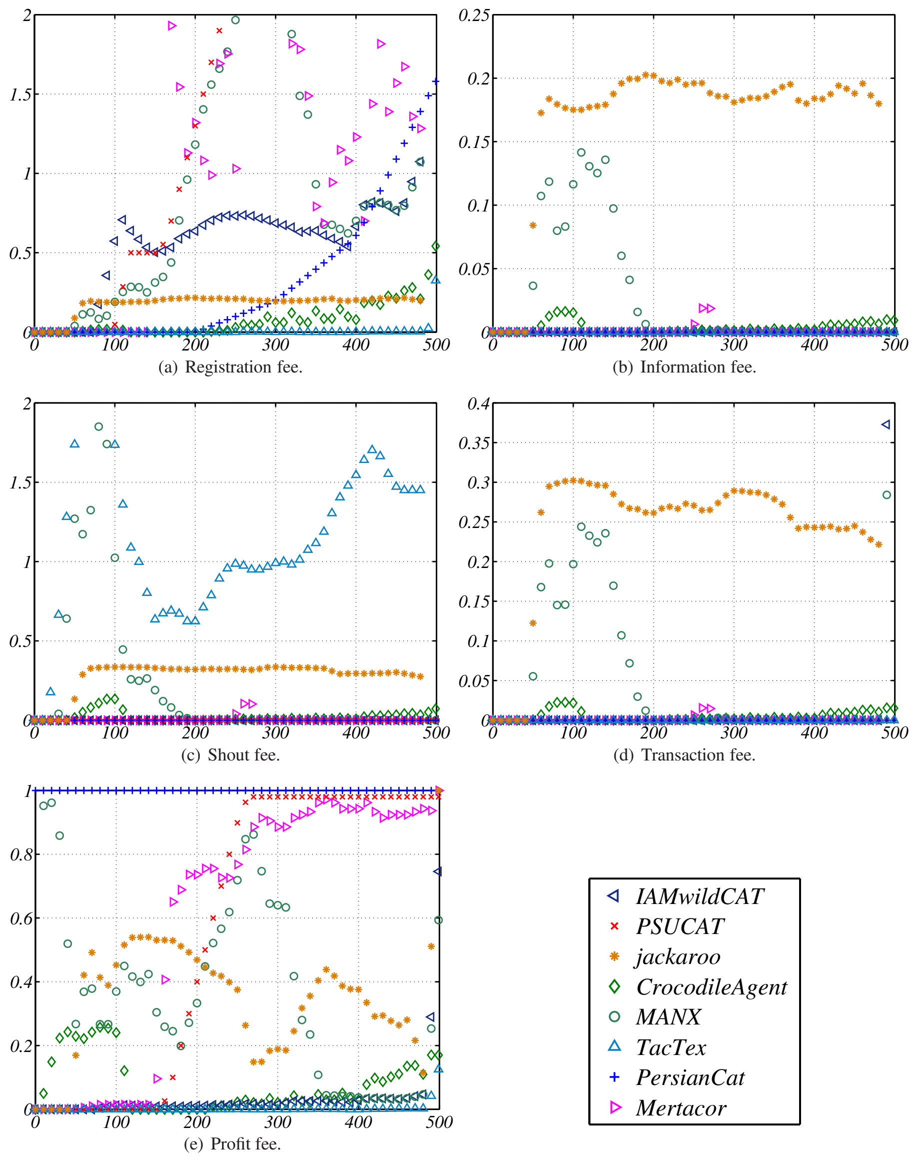
Fig. 2: Scores of specialists in our experiments. For key, see Fig 3. In all figures, the x-axis displays the number of trading days, and the y-axis gives the relevant score for each trader.

sections were averaged over a total of ten games and each datum is the average of a ten-day sliding window around it.

The scores obtained by specialists in our experiments (Table 2) broadly agree with the rankings in the tournament [45]. The 2007 CAT champion, `IAMwildCAT`, scores highest in our experiments and `PSUCAT`, which placed second in the competition, comes second. The only changes in ranking are due to `TacTex` and `MANX` increasing their scores since they could fully participate in every game. Fig. 2 shows the daily components of the scores and Fig. 3 shows the daily charges made.

5.2 Trader migration

The competition among specialists is reflected directly by the migration of intra-marginal traders and extra-marginal traders. Traders migrate based on estimates of expected profits, where the estimate for a given specialist is based on past experience with that specialist. Generally speaking, the more intra-marginal traders and the fewer extra-marginal traders in a market, the more potential profit there is, and the easier it is to make transactions and achieve a high transaction success rate. To measure the balance of intra-marginal and extra-



(a) Registration fee.

(b) Information fee.

(c) Shout fee.

(d) Transaction fee.

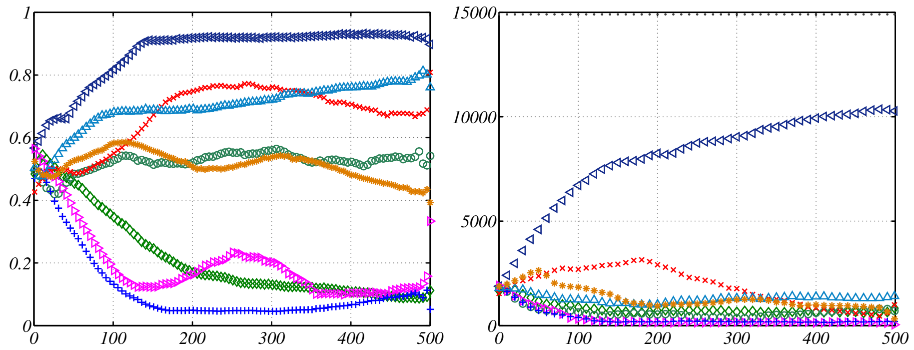
(e) Profit fee.

Fig. 3: Daily fees charged by specialists in our experiments. In all figures the x-axis displays the number of trading days.

marginal demand and supply, we introduce the *marginal coefficient*, β. For demand,

$$\beta_D = \frac{D_i}{D_i + D_e} \tag{3}$$

where D_i is the intra-marginal demand — the equilibrium — and D_e is the extra-marginal demand. The marginal coefficient of supply, β_S, can be defined similarly. β_D varies between



(a) Daily marginal coefficient of demand, β_D.
(b) Daily equilibrium profit.

Fig. 4: Properties of daily equilibria for individual specialists. For key, see Fig 3. In both figures, the x-axis displays the number of trading days.

0 and 1. A value of 0 indicates that all the buyers in the market are extra-marginal while 1 indicates that all the buyers are intra-marginal. Fig. 4(a) shows the daily value of β_D for the specialists. Since β_D provides no information on the absolute equilibrium quantity or profit, Fig. 4(b) gives the daily equilibrium profits in these markets.

As Fig. 4(a) shows, $\beta_D \approx 0.5$ in all the markets when the game starts. Then β_D of `IAMwildCAT` `TacTex`, and `PSUCAT` increases while that of `CrocodileAgent`, `PersianCat`, and `Mertacor` decreases. Since a falling β_D indicates losing intra-marginal traders and/or gaining extra-marginal traders, these changes indicate that intra-marginal traders and extra-marginal traders have different preferences over the different markets.

Intra-marginal traders seem to be sensitive to matching policies and charges, especially charges on profit. However, they seem to be relatively insensitive to other charges as long as they can still profit from trades. Fig. 4(a) shows that β_D of `Mertacor`, `PersianCat`, and `CrocodileAgent` decreases significantly at the beginning of the game and remains low all the way through the game. However these decreases occur for different reasons.

The low allocative efficiency of `Mertacor`, shown in Fig. 5, means a great portion of the potential social welfare is not achieved, suggesting an inefficient matching policy.[11] A close examination of `Mertacor`'s mechanism found that its MV-like matching policy strategically executes extra-marginal trades so as to increase its transaction success rate, but this leads to much lower profit for intra-marginal traders involved in those trades. In addition, `Mertacor` disregards the unmatched shouts every time the market is cleared. The traders that make these shouts are then unable to either improve their standing shouts or place new ones since the game server believes they still have active shouts. Some of these traders may be intra-marginal traders, causing unrealized intra-marginal trades. These two issues provide sufficient reason for intra-marginal traders to flee.

`PersianCat` and `CrocodileAgent` both lose traders due to imposing high profit charges. `PersianCat` charges 100% on profit for the whole game, as shown in Fig. 3(e), and this drives β_D down very quickly. `CrocodileAgent` levies a lower fee than `PersianCat` and therefore has

[11] During the CAT 2007 competition, some specialists announced invalid fees on some trading days, causing them to be banned from the games for a certain period. This is equivalent to the use of a very inefficient matching policy. Our experiments rounded their fees into the valid ranges and avoided banning the specialists.

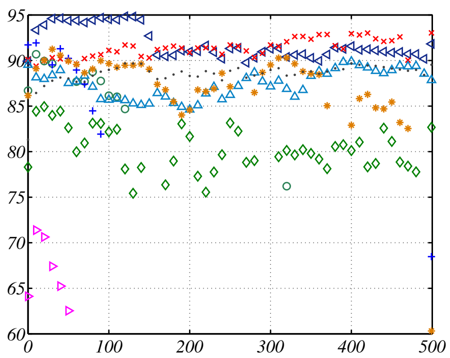


Fig. 5: Daily allocative efficiency in the markets. For key, see Fig 3. The x-axis displays the number of trading days.

a modestly decreasing β_D as shown in Fig. 4(a). The decrease of β_D in PSUCAT and jackaroo starting from days 250-300 follows the aggressive increase in the profit fee.

The rest of the specialists have much higher β_D despite their use of similar policies. IAMwildCAT, for instance, though adopting a version of MV, refrains from using it in the early rounds of a day, which usually are sufficient to realize most intra-marginal trades. MANX, though levying a high, yet volatile, profit fee, also levies other fees without bias considerations, which together scare away both extra-marginal traders and intra-marginal traders at an approximately same pace. Its β_D therefore zigzags around 0.5. The three specialists that obtain a β_D higher than 0.6 during the most time of the game, IAMwildCAT, PSUCAT, and TacTex, all produce allocative efficiency higher than 85%, again suggesting the importance of matching policies in keeping a high-quality trader population.

Registration fees appear to help to filter out extra-marginal traders, and information fees have the same effect on GD and ZIP traders (which require such information). Figs. 3(a) and 3(b) show that IAMwildCAT and jackaroo constantly impose one or both of these fees. As a result, the numbers of extra-marginal traders in those markets falls the most (see Fig. 6).

Shout fees also affect extra-marginal traders, but the degree of the effect depends on the shout accepting policy used. If the accepting policy is a strong filter and extra-marginal traders have little chance to place shouts, they can avoid losing money due to charges and thus are indifferent to shout charges. Their staying with a specialist therefore does not harm to the market's transaction success rate, and on the contrary, only adds to its market share. TacTex, uniquely among the specialists, charges only shout fees and consistently does so all the way through the game, as shown in Fig. 3(c). This policy together with its AA accepting policy — the weakest one possible — causes the extra-marginal traders to leave quickly as Fig. 6 demonstrates.

Mertacor managed to attract a lot of extra-marginal traders during the first 200 days, as shown in Fig. 6, due to its policy of not charging. Its policy change, starting to charge heavily on registration as in Fig. 3(a), explains why it loses almost all its extra-marginal traders shortly afterwards and its β_D increases significantly around day 200. Actually, higher registration fees in PSUCAT after day 150 and PersianCat after day 200, are both accompanied with a loss of market share in extra-marginal traders. CrocodileAgent increases its registration fee as well around day 200 but the modestly increased fee is still lower than those charged by most of other specialists, therefore it is still popular among extra-marginal traders.

In conclusion, extra-marginal traders, as expected, flee from those markets with high registration fees and information fees (and high shout fees in TacTex) to other markets, while intra-marginal traders migrate from markets with high profit fees and inefficient matching policies to those that do not have high charges and realize the most potential social welfare.

5.3 Learning and adaptation in specialists

The numbers of traders registered daily with the specialists, the profit made in the markets, and the daily charges made by the markets are all accessible to specialists via CATP. This makes it possible for specialists to learn and adapt their own policies. The transaction success rates however are unavailable unless a specialist is willing to obtain shout and transaction information directly from other specialists, paying any necessary fees. Specialists' payments for this purpose are not observed during the games.

Though specialists may adapt various types of auction policies, changes in charging policies are more obvious than other aspects from the data collected. MANX copies the charges of the leading markets in terms of profit share and market share combined, producing the most scattered charges among the specialists through the games. Looking at its charges gives us an approximate pattern of adaption of the other markets:

1. At the start, PersianCat charges the most (though only profit fees) while most of the others do not charge.
2. TacTex then starts to impose shout fees, but its payoff and winning position is not sustainable. Its market share declines significantly as seen in Fig. 2(b) around day 20.
3. Around day 50, jackaroo begins to impose heavy fees of all types, and like TacTex, jackaroo's market share decreases. Fig. 2(b) shows that before day 50, jackaroo attracts more and more traders, but after that, traders flee, quickly at first and then more slowly. Figs. 4(a), 4(b), and 2(b) further indicate that intra-marginal traders are more sensitive and flee faster than extra-marginal traders immediately after day 50, causing a plunge in market share immediately after day 50 and an increasing β_D between days 50 and 100. Around day 100, β_D starts to drop as well, suggesting extra-marginal traders leave at a slower and slower pace and intra-marginal traders continue to leave.
4. From around day 85, IAMwildCAT, which had previously not charged, starts to charge registration fees, as shown in Fig. 3(a), which scares away extra-marginal traders, and Fig. 4(a) shows a significantly faster increase of β_D. PSUCAT later does the same thing and causes an increasing β_D before days 100 and 150.[12]
5. IAMwildCAT and jackaroo, are designed to take advantage of the known length of games. They both increase their charges to much higher levels and make huge profits during the last days of the games, though JCAT takes measures to avoid traders going bankrupt in this situation and disregards any charges that traders cannot pay. The huge daily profits obtained, however, did not greatly increase the final scores of these specialists since the scoring mechanism adopted by CAT normalizes profits before scoring.

[12] The y axis in Fig. 3(a) has an upper bound of 2, and does not show the constant registration charges of 10 made by PSUCAT in the second half of the game. We do this to obtain a better general view, avoiding the chargs of other specialists (usually below 2) being squeezed together and becoming unreadable. The even higher charges by the specialists near the end of the game are not shown in Figs. 3(a)-3(d) for the same reason.

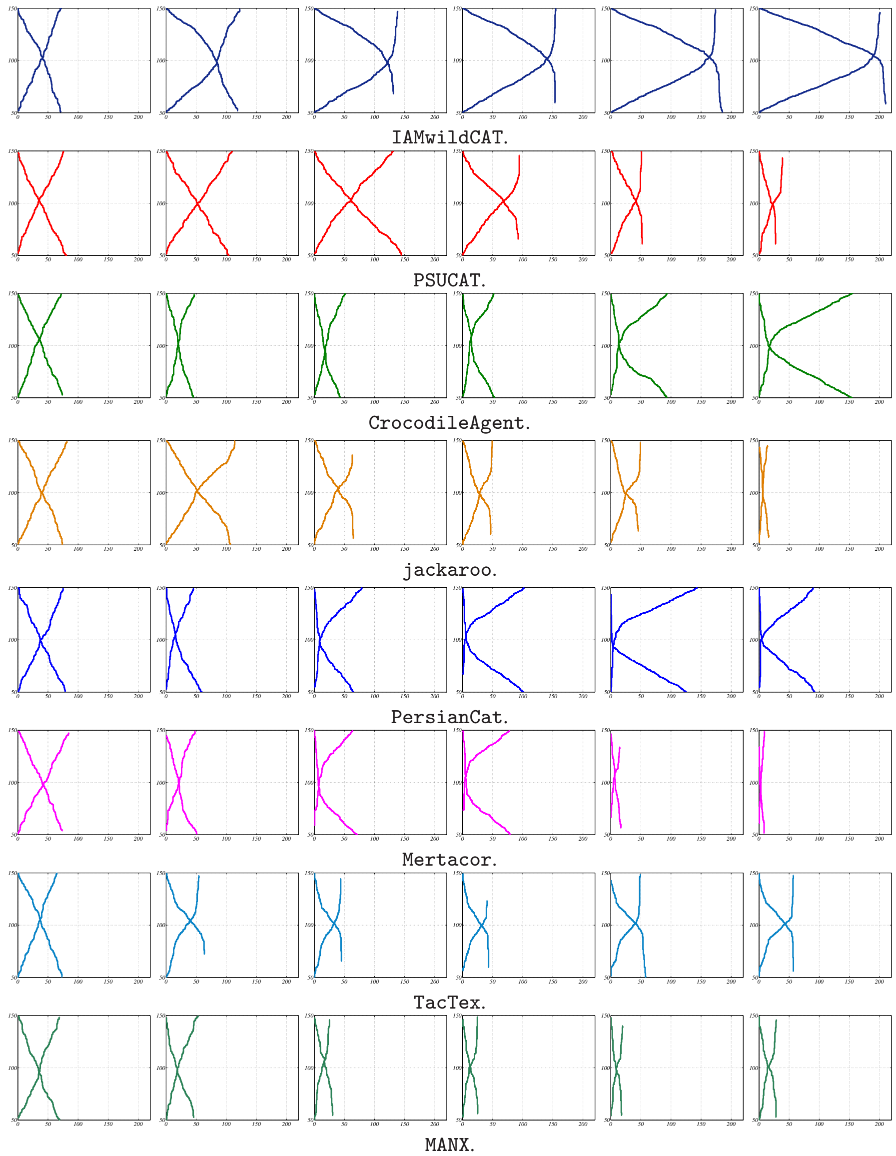


Fig. 6: Supply and demand curves for individual markets over time. Each graph has quantity on the x-axis and price on the y-axis. The leftmost graph gives supply and demand on day 0, and the remaining graphs in each row are those from days 50, 100, 150, 300, and 499 respectively. These graphs are from the same single run of the game.

The comparison between the charges of MANX, which copies charges, and those of the specialists mentioned above clearly shows which have adapted their policies and become the daily front-runners at each point.

IAMwildCAT exhibits stable performance according to almost all criteria and is worth further investigation. Profit share is the most sensitive metric since fee changes may immediately and dramatically cause the relative profit shares to go up or down. In Fig. 2(c), TacTex, jackaroo, and PSUCAT, one after another, increase their charges and claim big profit shares. However every subsequent increase leads to an apparent profit share drop for the previous front-runner, including what IAMwildCAT does to PSUCAT by increasing its profit charge gradually as shown in Fig. 3(e).[13] Despite this common theme, IAMwildCAT is to a great extent immune to the changes of other specialists' charges in terms of its profit share. This should be attributed to its target-oriented charging policy and the direct calculation of fees to achieve a certain target profit. Mertacor takes a similar approach, but its sub-optimal calculation method and other problematic auction rules prevent the approach from working well.

5.4 Discussion

Here we extract some general guidance for market design from the analysis above.

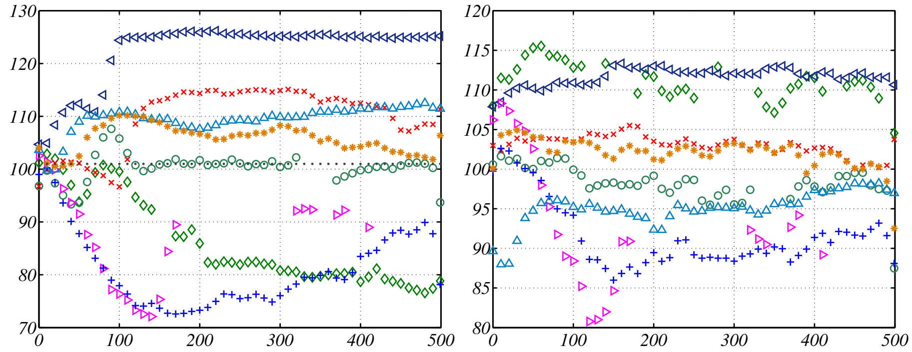
5.4.1 ME versus MV

If a high transaction success rate is desirable, then specialists have to explicitly take this into account, for example by matching intra-marginal and extra-marginal shouts, as MV does. However, caution should exercised when using an MV-like policy.

MV may cause intra-marginal traders to lose profits and in a competitive situation may lead them to prefer non-MV markets. This is exactly what happened to Mertacor. In addition, the extra-marginal trades may lower market efficiency. IAMwildCAT's matching policy is a mixture of ME and an MV-like policy. It uses the former in the first few rounds and the latter in the rest of the day. Fig. 2(d) shows that IAMwildCAT obtains high transaction success rates, very close or equal to 100%, after day 150 when the specialist starts to use the MV-like policy for more rounds in a day. As a consequence, IAMwildCAT's efficiency has a striking 5% drop, as shown in Fig. 5. Unlike Mertacor, IAMwildCAT did not show a loss of intra-marginal traders when it did this. This is because most of the intra-marginal traders traded in the early rounds of each day — when the MV-like policy was used, most of the traders still shouting were extra-marginal traders, few shouts made by these traders can pass the specialist's shout accepting policy, and these limited extra-marginal shouts did no great harm to the remaining intra-marginal traders.

Since traders are profit-seeking, MV-like policies can actually increase market allocative efficiency in some cases. For instance, a greedy intra-marginal trader may make an extra-marginal shout, which, when ME is used, will not be matched and therefore add to the number of unrealized intra-marginal trades. When MV is used, this extra-marginal shout can be matched by an intra-marginal trader, and the efficiency loss can thus be reduced or avoided. However, as [9] point out, such a matching policy may not gain much in volume and may be much less efficient.

[13] The increase of shout fees in TacTex around day 300 may also play a role in lowering PSUCAT's profit share.



(a) Underlying demand, as calculated from trader private values.

(b) Reported demand, as calculated from shouts

Fig. 7: Daily mean demand prices. For key, see Fig 3. The x-axis displays the number of trading days.

5.4.2 *Open versus closed shout accepting*

Shout accepting policies have a direct impact on the effectiveness of other auction rules. An open shout accepting policy places a heavy burden on the matching policy. When the matching policy is also ineffective, intra-marginal traders fail to profit and tend to leave. In contrast, if the shout accepting policy filters out most extra-marginal shouts, a simple matching policy can work well.

For example, `CrocodileAgent` and `PersianCat` have similar trader populations in terms of competitiveness as shown in Figs. 4(a), 6, and 7(a), and they both use the ME matching policy. However, they produce significantly different shout sets as shown in Fig. 7(b) and transaction success rates as illustrated in Fig. 2(d). This is due to the AE accepting policy in `CrocodileAgent`, which is much more effective than the policy in `PersianCat`.

In addition, AQ, the common shout accepting policy, may leave the door wide-open at the start of days. In CAT games, shouts automatically expire at the end of a day. This resets the market quotes in AQ and loses valuable information from the previous day on the underlying demand and supply schedules, which do not usually change dramatically over days. This may explain why `jackaroo` and `MANX`, the two AQ markets, with higher mean theoretical demand prices in Fig. 7(a) and better shaped demand and supply as shown in Fig. 6 than those in `CrocodileAgent`, produce bid sets with lower mean prices as in Fig. 7(b) and lower transaction success rates.

We believe a good shout accepting policy in the current CAT game setting should be able to reflect the collective properties of traders and carry this knowledge from day to day, as the history-based policy AH does. We expect most specialists would be better off using AH, and later in the paper we present some experimental evidence that suports this suggestion. `PSUCAT`'s customized AE is another potential policy. The mean theoretical demand price in the `PSUCAT` market jumps around day 100 in Fig. 7(a), and β_D follows in Fig. 4(a), but this did not cause the mean bid price in Fig. 7(b) to climb as well, indicating the effectiveness of its shout accepting policy, which successfully prevented extra-marginal traders placing shouts.

5.4.3 Market share versus profits

In CAT games it is common for specialists to find that increasing fees can boost profits but gradually lead to loss of market share. If market share falls too low, however, such profits cannot be sustained. In contrast, low charges help to gain market share but harm profits. If a charging policy is properly designed, it may keep both measures at suitably high levels. Imposing small, flat, fees, after a game has been running for a while, may not have much negative effect on market shares if the good reputation of a specialist has been established and the traders continue to make a profit that is much higher than the fees. In this way, on the basis of a big market share, small fees may still bring a considerable amount of profit. `IAMwildCAT` demonstrates this.

Bias towards different types of fees in charging policies can also benefit specialists. For example, `IAMwildCAT` and `PSUCAT` use registration fees to drive extra-marginal traders away. Reducing the number of extra-marginal traders makes it easy for the remaining trader population to find partners, and for the specialists to obtain high transaction success rates. However as discussed in Section 5.2, a powerful shout accepting policy may make this unnecessary or even harmful, since such a policy may filter out most extra-marginal shouts and avoid their negative effect on transaction success rates. With a strong shout accepting policy and without charges on registration and information, a market actually becomes a free place for extra-marginal traders to stay. If other markets impose these charges, these traders are sure to be willing to stay with a market that doesn't charge, and hence boost market share.

5.4.4 Targeted versus non-targeted charges

Specialists in the competition adapt their daily charges differently, as shown in Table 3. Some do this by setting specific performance targets, determining these targets from estimates of the expected actions of other specialists, while others increase or decrease their current charges without setting targets or modeling the effect of the changes. `IAMwildCAT`, for instance, determines a reasonable portion of the profit it desires to make via registration fees, and calculates its registration fee and profit fee by taking into consideration the average profit a trader has been able to make in its market. In contrast, parameter values and charge levels of most other specialists are decided rather arbitrarily. As a result, `IAMwildCAT` has a stable performance in the face of changes by other specialists.

Several specialists are reactive, copying the fees that other, well-performing, specialists charge. `MANX` in particular does this. This approach is problematic for two reasons. First, it is usually based on a short-term assessment and may not optimize the long-term outcome. Second, copying a winning specialist may not be a winning strategy. The effect of fees is closely linked to the other auction rules that specialists adopt and the properties of their trader population at that moment [26]. `MANX`'s follow-the-leader approach demonstrates impressive performance during the early part of a game when the trader populations in all individual markets are quite similar. However it fails to lead to a similar outcome after traders have converged to prefer different markets.

6 A black-box analysis of CAT 2007 entries

The above white-box analysis is feasible only when the internal structure of each specialist is known, and can only be conducted in very limited situations because it requires a thorough manual examination of game dynamics. A black-box analysis abstracts away the internal structure of the specialists and many details of the dynamics during the interaction between specialists, making it possible to consider many more situations. However, an exhaustive black-box analysis may still involve high complexity. This is due to the fact that a game may have an arbitrary number of entrants and an arbitrary number of specialists.[14] The results of n-entrant, m-specialist games may not necessarily agree with the results of $(n+1)$-entrant, m-specialist games, or n-entrant, $(m+1)$-specialist games. For instance, entrant A beating B in a bilateral game does not necessarily imply that A would still beat B when an additional entrant C joins the game, no matter whether C uses either of the specialists used by A and B, or a third, new specialist. This suggests, for example, that the replicator dynamics fields reported in [32] based on 6-agent auction games or in [16] based on 6-agent TAC SCM games are likely to change when a different set of game profiles are used.

To further explore the CAT 2007 entries, we ran two more sets of experiments — multi-lateral simulations with games involving all the specialists and bilateral simulations with games each involving two specialists. In some of these experiments, we considered an additional specialist, `MetroCat`. `MetroCat` is a CDA market using the AH shout accepting policy, and a detailed description of the implementation is provided in Appendix A. Since we developed the competition platform, `MetroCat` was not an entry in the competition, but it was included in the JCAT source code provided to entrants to the 2007 Market Design game to support the development of their specialists (rather as Axelrod provided a description of "Tit for Two Tats" to entrants in the initial Prisoner's Dilemma tournament [1]). Here we use `MetroCat` as a benchmark in our post-tournament experiments. Note that the inclusion of `MetroCat` is the only substantive deviation that our experiments make from the format of the Market Design game as it ran in 2007. We do not, for example, consider more entrants than the eight considered above (aside from introducing `MetroCat`) and we concentrate on situations in which each entrant uses a different specialist. This is because our aim here is to learn more about the competition rather than to perform an exhaustive analysis of ways the competition might have unfolded and because an exhaustive analysis was not feasible since each game runs for around five hours. [15] A full understanding of the CAT game would, however, require such an exhaustive analysis.

6.1 Multi-lateral simulations

Inspired by ecological analyses like [1, 37] — in which more copies of successful specialists, and less copies of unsuccessful specialists are run for each successive game — but constrained by the number of specialists that we could have in a single game, we ran the following experiment. One copy of each of the specialist from the 2007 CAT competition was run for the full five hundred days of the game. The specialists in this game we considered to be the first "generation" of the analysis. We then ran a second game, with a second generation of specialists. This second generation still contained one copy of each specialist, but each was run for the fraction of the 500 trading days proportional to the score of that specialist in the first generation (as a fraction of the total score). A third generation was then

[14] Each entrant has a choice of developing a new specialist or reusing one from the repository so there could be less distinct specialists than entrants.

[15] Irrespective of the hardware — the length of each trading day is hard-coded at a constant that permits each specialist to take time to perform possibly complex computations — any reduction in this time would potentially distort the results by preventing some specialists from performing as designed.

run in which each specialist was run for a number of trading days that matched the score that specialist obtained in the second experiment, and so on.

Figs. 8(a) and 8(b) show the result of this simulation. The distribution on the y-axis shows the proportion of the total number of trading days for all markets that are allotted to each market, indicating how this evolves in populations without and with `MetroCat` respectively. Fig. 8(a) shows that without `MetroCat`:

- the results of this analysis agree with the results reported in [23], again confirming that `IAMwildCAT` was the strongest entry in the 2007 competition; and
- the days allotted to `PersianCat` shrink more slowly than those allotted to other losing specialists. This agrees with the results of bilateral games between `IAMwildCAT` and `PersianCat` (described below) and suggests that `PersianCat` was a strong entry, stronger than its overall position suggests.

Fig. 8(b) shows that with `MetroCat`:

- `MetroCat` quickly dominates the other entries, doing so faster than `IAMwildCAT` in Fig. 8(a), so that by generation 8 only `MetroCat` has any trading days; and
- the CAT 2007 champion, `IAMwildCAT`, loses trading days faster than other entries after generation 1, indicating some weakness in its design when facing an opponent like `MetroCat`.

In both cases, one specialist quickly comes to dominate the others.

6.2 Bilateral simulation

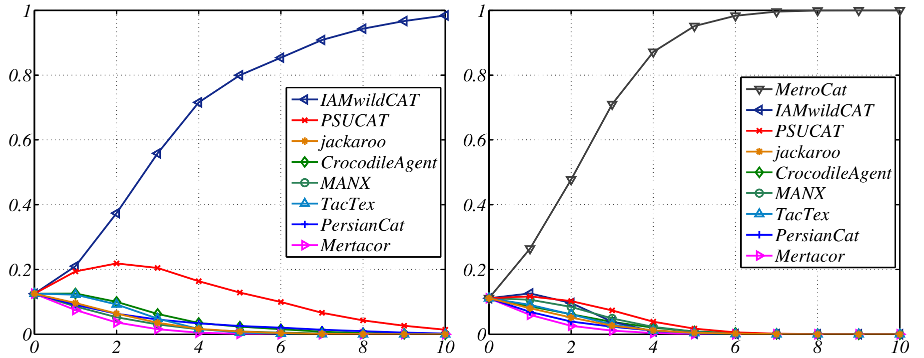
One-on-one games allow us to examine in detail the strength and weakness of each specialist when it faces different opponents.

6.2.1 Payoff table

We ran 81 one-on-one games, that is one for every pair of the nine specialists we have been considering, including `MetroCat` and including nine self-play games. Table 4 shows the resulting payoffs of specialists — their average daily scores — in these CAT games. Each payoff is averaged over ten iterations and entry (i, j) is the payoff of specialist i in the game against specialist j. Thus in a game between `MetroCat` and `IAMwildCAT`, `MetroCat` scores 0.71 and `IAMwildCAT` scores 0.59.

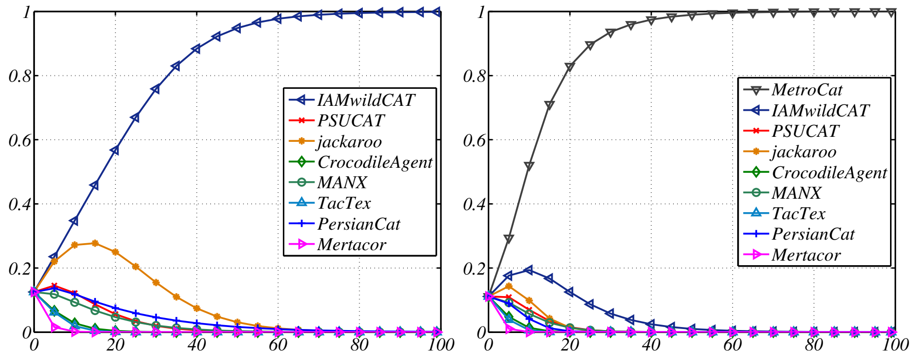
Fig. 10 compares these payoffs pictorially using a polar coordinate system. Each plot shows the nine specialists evenly distributed on the outer circle, the radial coordinates of the nine vertices of the solid-line polygon represent a given specialist's payoffs against all nine specialists, and the radial coordinates of the nine vertices of the dashed-line polygon represent its opponents' payoffs in these games. The solid-line polygon and the dashed-line polygon overlap on the vertex that corresponds to the self-play game of the particular specialist. For example, in Fig. 10(a), the solid-line polygon completely encloses the dashed-line one, meaning that `MetroCat` scores more than all the other specialists in bilateral competitions.[16] In Fig. 10(i) the solid polygon lies within the dashed polygon showing that `Mertacor` loses every bilateral game. The two polygons for all other specialists intersect, indicating a more complex relationship between them.

[16] `MetroCat` maintains a better balance than those CAT 2007 entries between market share and profit share by keeping extra-marginal traders and preventing them from placing uncompetitive shouts.



(a) Without `MetroCat` (b) Including `MetroCat`

Fig. 8: Ecological simulation of CAT 2007 entries based on multi-lateral CAT games. In both graphs the x-axis gives the number of generations, and the y-axis gives the proportion of each kind of specialist.



(a) Without `MetroCat` (b) Including `MetroCat`

Fig. 9: Ecological simulation of CAT 2007 entries based on bilateral CAT games. In both graphs the x-axis gives the number of generations, and the y-axis gives the proportion of each kind of specialist.

Both Figs. 10(b) and 10(h) show that `IAMwildCAT`, the CAT 2007 champion, surprisingly loses, albeit narrowly, against `PersianCat`, which placed sixth in the competition. This provides an explanation for the fact that in Fig. 8(a) the days for `PersianCat` shrink more slowly than those for other specialists — it does well against the increasingly dominant `IAMwildCAT`. `IAMwildCAT` losing to `PersianCat` along with the defeat of `PersianCat` by `PSUCAT` and `jackaroo`, suggests that `IAMwildCAT` has some particular weakness that is taken advantage of by `PersianCat`.

Other discrepancies, when compared to the results of the 2007 competition, include `jackaroo` (which placed fourth) winning over `PSUCAT` (second) and `Crocodile` (third). These may be significant, or may be caused by differences in the configurations for `PSUCAT` and `Crocodile` used in the simulations and CAT 2007 games.

Table 4: The payoff matrix of bilateral CAT games between CAT 2007 entries and `MetroCat`. Each row gives the average daily score for that specialist in games against all the other specialists, and in self-play.

specialist	Metro	IAM	PSU	jack	Croc	MANX	Tac	Pers	Mert
MetroCat	0.65	0.71	0.75	0.78	0.82	0.75	0.86	0.78	0.89
IAMwildCAT	0.59	0.66	0.72	0.68	0.77	0.71	0.80	0.61	0.76
PSUCAT	0.54	0.57	0.62	0.55	0.70	0.61	0.64	0.74	0.83
jackaroo	0.48	0.59	0.70	0.63	0.75	0.71	0.78	0.69	0.86
CrocodileAgent	0.44	0.52	0.54	0.51	0.49	0.46	0.62	0.59	0.73
MANX	0.54	0.59	0.61	0.58	0.51	0.64	0.72	0.62	0.69
TacTex	0.34	0.41	0.57	0.43	0.63	0.54	0.55	0.61	0.72
PersianCat	0.43	0.62	0.52	0.59	0.70	0.67	0.64	0.64	0.77
Mertacor	0.27	0.38	0.29	0.32	0.51	0.40	0.45	0.47	0.55

6.2.2 Ecological simulation

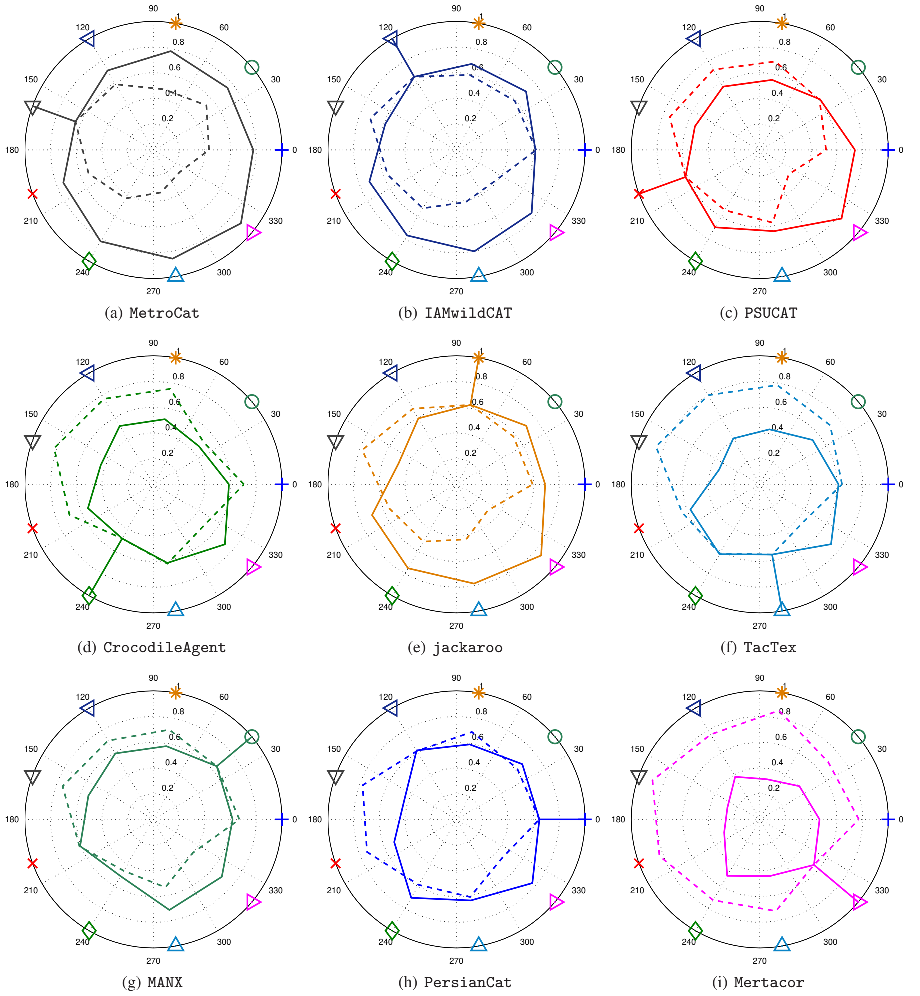
The payoff table for the bilateral CAT games can be used to approximate ecological dynamics for populations involving more than two specialist types. The payoff of each specialist type for a certain population mixture is computed as the expected payoff for this specialist assuming that each specialist obtains the payoff it would have obtained had it computed one-on-one with each of the other specialists in the mix. Under this assumption, Figs. 9(a) and 9(b) show how a population that starts with an even distribution of specialists evolves over time when, as in [1], every specialist plays against every other specialist in every generation in bilateral games, and the number of specialists in any generation is proportional to the payoff achieved by that "breed" of specialist in the previous generation.

Comparing Fig. 9(a) with Fig. 8(a), and Fig. 9(b) with Fig. 8(b), shows that while the winning strategies are the same, the ecological simulations based on multi-lateral games converge much faster than those based on bilateral games (the scales on the x-axis are very different in the two sets of plots). This may be explained by the fact that bilateral games give strategies a chance to benefit from the ability to perform well against specific opponents, whereas in the multi-lateral games they have to be good against all opponents in oder to survive. Another noticeable phenomenon is that `PSUCAT` performs much worse in the simulations with bilateral games than those with multi-lateral games, while `jackaroo` and `IAMwildCAT` do the opposite. These discrepancies indicate that, as one might expect, different game setups may lead to very different results. However, our results may be helpful to identify the weakness in strategies by looking at the particular scenario in which a strategy performs poorly.

6.2.3 Offense, defense, dominance, and equilibrium

To further reveal the strength and weakness of specialists, we compare specialists' payoffs — which we call *offense* — and the payoffs they allow opponents to make — which we call *defense* — when they face a same opponent.

Fig. 11(a) puts all the solid-line, offense, polygons in Fig. 10 into a single polar coordinate system and Fig. 11(b) shows all the dashed-line, defense, polygons in a similar way. The color scheme is the same as in Fig. 10. The comparison shows clearly that `MetroCat`



(a) `MetroCat` (b) `IAMwildCAT` (c) `PSUCAT`

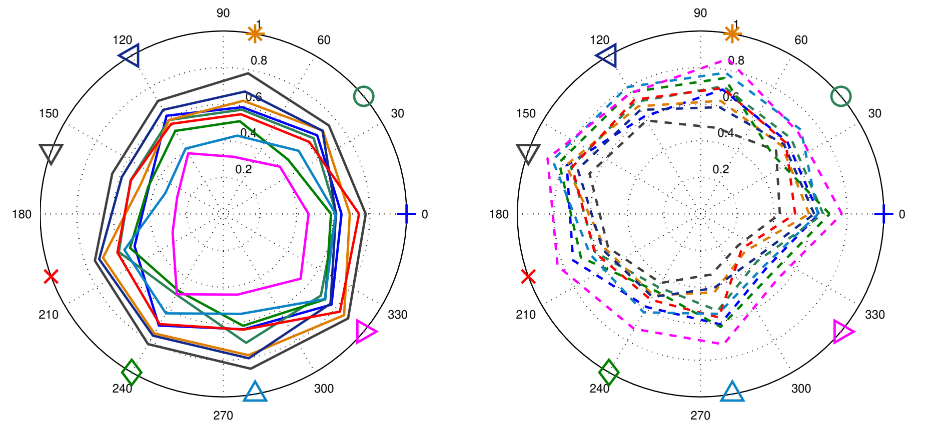
(d) `CrocodileAgent` (e) `jackaroo` (f) `TacTex`

(g) `MANX` (h) `PersianCat` (i) `Mertacor`

Fig. 10: Payoffs of self and opponents in bilateral CAT games. On the outer circles starting from polar angle 0° lists the nine specialists anti-clockwise : `PersianCat` (0°), `MANX` (40°), `jackaroo` (80°), `IAMwildCAT` (120°), `MetroCat` (160°), `PSUCAT` (200°), `CrocodileAgent` (240°), `TacTex` (280°), and `Mertacor` (320°). The radial coordinates of the nine vertices of the solid-line polygon represent a given specialist's payoffs against all nine specialists respectively, and those of the dashed-line polygon represent payoffs of its opponents. The overlapping vertex of the two polygons in each plot is the self-play game of the particular specialist.

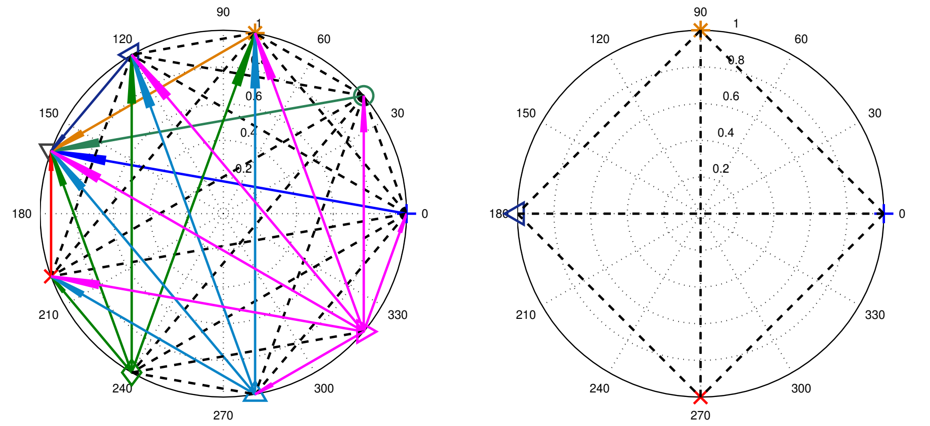
has both the strongest offense and the strongest defense, while `Mertacor` exhibits almost the opposite.

By analogy with a two-player normal-form game in which both players choose to play like one of the nine specialists, we say a specialist *dominates* another if the offense of the



(a) Offense (b) Defense

Fig. 11: Comparison of the offense and defense of specialists. For keys, see Fig. 10.



(a) S_6,S_3,S_8,S_2,S_9,S_5,S_1,S_7,S_4 anti-clockwise from 0° (b) S_6,S_8,S_2,S_5 anti-clockwise from 0°

Fig. 12: Dominance relations based on offense. S_1: `CrocodileAgent`, S_2: `IAMwildCAT`, S_3: `MANX`, S_4: `Mertacor`, S_5: `PSUCAT`, S_6: `PersianCat`, S_7: `TacTex`, S_8: `jackaroo`, and S_9: `MetroCat`.

former is better than the offense of the latter for every opponent that they face.[17] Considering the results of the bilateral games we can analyze them for such dominance relationships, and the result is shown in Fig. 12. This represents each dominance relation with an arrow starting from the dominated specialist to the dominating one, and the unavailability of a dominance relation between two specialists with a dashed line. Our goal here, as is usual in normal-form games, is less to identify the an overall dominant specialist than to reduce the number of specialists we consider by iteratively removing dominated specialists. We can then investigate the relative strength and weakness of the undominated specialists at a

[17] Defense may also be used to define dominance but we do not investigate this here.

lower computational cost.[18] Fig. 12(a) shows, as Fig. 11 already does, that `MetroCat` dominates all the other specialists while `Mertacor` is almost dominated by all the rest except for `CrocodileAgent`. If we eliminate the dominating `MetroCat` (which, of course, was not in the 2007 competition) and any specialist that is dominated by at least one specialist other than `MetroCat`, we end up with `PersianCat`, `IAMwildCAT`, `PSUCAT`, and `jackaroo`. We are thus left with Fig. 12(b) which shows the relationships between the offenses of the best specialists in the 2007 competition as judged by bilateral games.

For each set of three specialists from the four, we carried out a heuristic strategy analysis like that in [32,48], and Fig. 13 shows the four replicator dynamics fields that result. To create these plots, we used Table 4 to compute the payoff for each specialist type in a certain population mixture in the same way as in the ecological simulation above. Figs. 13(a) and 13(c) show an unstable equilibrium between `IAMwildCAT` and `PersianCat`. Without considering this equilibrium and the pure profiles, all other profiles lead to a homogeneous population — `IAMwildCAT` in Figs. 13(a), 13(b), and 13(c), and `jackaroo` in Fig. 13(d). This indicates that `IAMwildCAT` comes close to dominating the winning strategies in these 3-specialist scenarios. This clear victory may be due to the youth of CAT tournaments (it was certainly the case that `IAMwildCAT` did not win in 2008). As CAT competitions continue, and the strategies for specialists evolve, we expect that the relative strength between these strategies would become more complex and some mixed equilibrium may start to emerge.

7 Summary and Conclusions

This paper reports a post-competition study of the entrants in the 2007 TAC Market Design, or CAT, competition. This work has made several novel contributions to the study of electronic markets, and to the design and analysis of market mechanisms.

First, this paper provides a more extensive assessment of the performance of the entrants to the 2007 competition than was possible in the competition itself. Each competition game ran for around 8 hours, and given the technical problems experienced by both organizers and competitors, this meant that it was only possible to run two games during the three days of the competition, and not all games involved all competitors. Running more games and including all the competitors gives more definitive results, and confirms the superiority of `IAMwildCAT` that was seen in the actual tournament. The various bilateral and multi-lateral simulations we have undertaken have also revealed weaknesses of some specialists in some situations, for instance, the defeat of `IAMwildCAT` by `PersianCat` in bilateral games, and the poorer relative performance of `jackaroo` in multi-lateral rather than in bilateral games.

Second, this paper provides the first classification of the strategies used by 2007 Market Design competition entrants, and the first comparison of the effects of these strategies in a rigorous, systematic experiment. While there are many more experiments to be run before we fully understand the comparative strengths of the strategies, we believe that these aspects of the paper will be of help to future entrants in the competition.

Third, the paper explores the implications of the design of the various components of double auction mechanisms, in particular the interaction between the component policies, and their effect on auction performance. We hope that this part of the paper will help to

[18] Imagine that a specialist may not be the best in a competition, but its designer may still want to improve it rather than simply adopting the winning strategy designed by others. This reduction based on dominance may help to zoom into those match-ups that are most worthy of examination, perhaps through a white-box analysis.

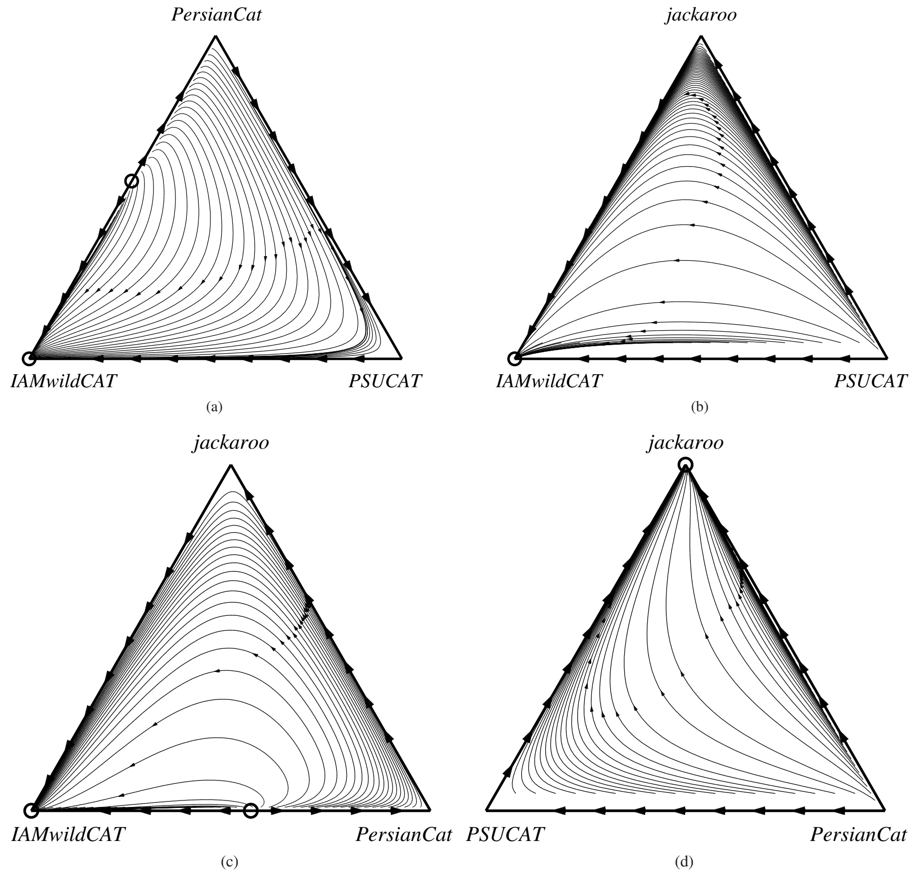


Fig. 13: Replicator dynamics fields for each set of three specialists from `PersianCat`, `jackaroo`, `IAMwildCAT`, and `PSUCAT`.

guide future research on the design of double auctions, not least in suggesting new market designs that involve new combinations of component policies.

Finally, we looked at the performance of the specialist `MetroCat`, which uses a history-based shout-accepting policy derived from the GD trading strategy. Running `MetroCat` against other specialists suggests that `MetroCat` would have done well had it been entered in the 2007 competition. This, in turn, suggests the importance of the specific shout-accepting policy that is at the heart of `MetroCat`, as well as indicating that the shout-accepting policy is an important consideration in the design of a market mechanism.

The success of such a simple specialist as `MetroCat` also suggests that there is significant room for improvement in the entries to the 2007 competition and this supported by other results that we do not have room to describe here (for example in some of the trial games, entrants struggled to beat classic double auctions with fixed charging policies). We have now had two further years of CAT competitions, and in the future we aim to analyse the improvements in specialist performance over those competitions, and to identify the causes of those improvements.

The bilateral games and multi-lateral games can be viewed as two ends of a spectrum of CAT games. The aim of running simulations based on both configurations is to explore whether the different competition configurations lead to different results. It is hoped that if they make no much difference — and our results suggest that they do not — then the low cost of bilateral games can be used to approximate the games involving more different individual types and different population distributions.

In related work, Kaisers *et al.* [18] explored the acquisition of the payoff table for n-player games based on the payoff table for 2-player games and *vice versa*, both involving a same set of strategies for players. They showed that the linear-programming-based approximation approach works well in games between trading strategies. Our simulation shows that an approximation approach may work but would need additional tuneup so as to reduce the distortions incurred. The discrepancies observed in this paper suggest that additional simulations may need to be run to obtain more accurate approximation. The problem they try to solve can actually be extended into a more general one: how to build (approximately) the payoff table for n-player games based on a set of complete or partial payoff tables each for games involving no more than n players. Suppose, in a n-player game, each player may choose one of s strategies. There are thus totally C^s_{n+s-1} possible match-ups. If each match-up is simulated, a heuristic payoff table would become available to generate a replicator dynamics field for the s strategies, where possible equilibria can be identified as well as the relative strength of each strategy. If the C^s_{n+s-1} match-ups be viewed as the C^s_{n+s-1} discrete points along the dimension denoted as $\mathscr{D}(n,s)$, the above problem becomes to run simulations for points scattering along lower dimensions, e.g., $\mathscr{D}(2,s)$ for 2-player games in [18], so as to approximate the results for the C^s_{n+s-1} points along $\mathscr{D}(n,s)$. This approach would have more flexibility and allow gradual distortion reduction over time.

This future work is desirable because, even if such an approximation is not quantitatively accurate, it may provide qualitative guidance on what scenarios should be investigated further, and help to reduce the overall computational complexity. For example, the replicator dynamics fields in both [16] and [34] suggest that if there is a mixed equilibrium between three strategies, there may be at least one mixed equilibrium between two of these strategies. In addition, for a space of heuristic strategies that still expands gradually, like that for CAT games, it is no less important to be creative — focusing on creating better strategies — than to be fair — finding a better way to evaluate existing strategies. Shedding light on the weaknesses of a strategy and directions to improve it may be even more important for strategies that are of practical importance.

Another possible line of future research is to consider the evolution of trading agents in addition to that of specialists. To make the situation simple, the simulations in this paper use a portfolio of trading agent strategies that does not evolve at all. It would be more realistic, however, to have simulations with intervening CAT competitions and trading agent competitions, so that trading agents learn to adapt their strategies as they interact with each other and with the specialists.

A further extension is to view a market mechanism as a combination of atomic auction rules rather than being an atomic entity itself. From this point of view, there would be multiple populations, each for a type of auction rule. Multiple individuals, one from each population, need to collaborate to form a complete market mechanism, which can then compete against other combinations. The payoff of a market mechanism from a simulation would be used as the payoff for each individual component of the market mechanism. This multi-population simulation may be considered as a *grey-box* approach, a mixture of black-box approach and white-box approach, since it considers the internal logic of specialist strategies. Such a grey-box approach can be used to explore a solution space, enabling an automated solution design method, as long as a modular strategy design is available. The parameterized framework for specialists presented in Section 3 forms an ideal foundation for further work along this line.

Acknowledgements The authors acknowledge financial support from the National Science Foundation under grant NSF IIS-0329037, *Tools and Techniques for Automated Mechanism Design*, and from the UK EPSRC under grant GR/T10657/01, *Market Based Control of Complex Computational Systems* and resources from the computational facility at the CUNY Graduate Center. Our grateful thanks go to the entrants to the 2007 Market Design game for releasing binaries of their specialists, to Haizheng Zhang for information on PSUCAT, to Elizabeth Sklar for advice on the design of JCAT, and to the reviewers of this paper for many helpful comments.

References

1. R. Axelrod. *The Evolution of Cooperation*. Basic Books, Perseus Books Group, New York, NY, December 2006.
2. T. N. Cason and D. Friedman. A comparison of market institutions. In *Handbook of Experimental Economics Results*, volume 1, chapter 33. North Holland, 2008.
3. S. H. Clearwater, editor. *Market-Based Control: A Paradigm for Distributed Resource Allocation*. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1996.
4. D. Cliff. Evolution of market mechanism through a continuous space of auction-types. Technical report, Hewlett-Packard Research Laboratories, Bristol, England, 2001.
5. D. Cliff and J. Bruten. Minimal-intelligence agents for bargaining behaviours in market-based environments. Technical report, Hewlett-Packard Research Laboratories, Bristol, England, 1997.
6. R. Das, J. E. Hanson, J. O. Kephart, and G. Tesauro. Agent-human interactions in the continuous double auction. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, Seattle, U.S.A., August 2001.
7. I. Erev and A. E. Roth. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *The American Economic Review*, 88(4):848–881, September 1998.
8. D. Friedman. The double auction institution: A survey. In Friedman and Rust [10], chapter 1, pages 3–25.
9. D. Friedman and C. Rich. The matching market institution. In *Handbook of Experimental Economics Results*, volume 1, chapter 13. North Holland, 2008.
10. D. Friedman and J. Rust, editors. *The Double Auction Market: Institutions, Theories and Evidence*. Santa Fe Institute Studies in the Sciences of Complexity. Westview Press, Perseus Books Group, 1993.
11. D. Fudenberg and D. K. Levine. *The Theory of Learning in Games*. MIT Press, Cambridge, MA, 1998.
12. E. Gerding, P. McBurney, J. Niu, S. Parsons, and S. Phelps. Overview of CAT: A market design competition. Technical Report ULCS-07-006, Department of Computer Science, University of Liverpool, Liverpool, UK, 2007. Version 1.1.
13. S. Gjerstad and J. Dickhaut. Price formation in double auctions. *Games and Economic Behavior*, 22:1–29, 1998.
14. D. K. Gode and S. Sunder. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1):119–137, 1993.
15. M. He, N. R. Jennings, and H.-F. Leung. On agent-mediated electronic commerce. *IEEE Transactions on Knowledge and Data Engineering*, 15(4):985–1003, 2003.
16. P. R. Jordan, C. Kiekintveld, and M. P. Wellman. Empirical game-theoretic analysis of the TAC supply chain game. In *Proceedings of the Sixth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1188–1195, Honolulu, Hawaii, May 2007.
17. P. R. Jordan and M. P. Wellman. Best-first search for approximate equilibria in empirical games. In *Proceedings of AAAI-07 Workshop on Trading Agent Design and Analysis (TADA-07)*, Vancouver, Canada, 2007.
18. M. Kaisers, K. Tuyls, F. Thuijsman, and S. Parsons. Auction analysis by normal form game approximation. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Sydney, Australia, December 2008. Short Paper.

19. P. Klemperer. How (not) to run auctions: The European 3G telecom auctions. *European Economic Review*, 46(4–5):829–845, 2002.
20. J. K. MacKie-Mason and M. P. Wellman. Automated markets and trading agents. In L. Tesfatsion and K. L. Judd, editors, *Handbook of Computational Economics*, volume 2, chapter 28, pages 1381–1431. Elsevier, 2006.
21. K. A. McCabe, S. J. Rassenti, and V. L. Smith. Designing a uniform price double auction. In Friedman and Rust [10], chapter 11, pages 307–332.
22. J. Niu, K. Cai, P. McBurney, and S. Parsons. An analysis of entries in the first TAC market design competition. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, Sydney, Australia, December 2008.
23. J. Niu, K. Cai, S. Parsons, E. Gerding, and P. McBurney. Characterizing effective auction mechanisms: Insights from the 2007 TAC Mechanism Design Competition. In Padgham, Parkes, Müller, and Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 1079–1086, Estoril, Portugal, May 2008.
24. J. Niu, K. Cai, S. Parsons, E. Gerding, P. McBurney, T. Moyaux, S. Phelps, and D. Shield. JCAT: A platform for the TAC Market Design Competition. In Padgham, Parkes, Müller, and Parsons, editors, *Proceedings of the Seventh International Conference on Autonomous Agents and Multiagent Systems*, pages 1649–1650, Estoril, Portugal, May 2008. Demo Paper.
25. J. Niu, K. Cai, S. Parsons, and E. Sklar. Reducing price fluctuation in continuous double auctions through pricing policy and shout improvement rule. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1143–1150, Hakodate, Japan, 2006.
26. J. Niu, K. Cai, S. Parsons, and E. Sklar. Some preliminary results on competition between markets for automated traders. In *Proceedings of AAAI-07 Workshop on Trading Agent Design and Analysis (TADA-07)*, Vancouver, Canada, July 2007.
27. J. Niu, A. Mmoloke, P. McBurney, and S. Parsons. CATP Specification: A communication protocol for CAT games. Technical report, Department of Computer Science, Graduate Center, City University of New York, New York, U.S.A., 2007.
28. D. Pardoe and P. Stone. Developing adaptive auction mechanisms. *ACM SIGecom Exchanges*, 5(3):1–10, April 2005.
29. A. Petric, V. Podobnik, A. Grguric, and M. Zemljic. Designing an effective e-market: an overview of the CAT agent. In *Proceedings of AAAI-08 Workshop on Trading Agent Design and Analysis (TADA-08)*, Chicago, IL, USA, 2008.
30. S. Phelps. JASA — Java Auction Simulation API. `http://jasa.sourceforget.net/`, 2005.
31. S. Phelps, M. Marcinkiewicz, S. Parsons, and P. McBurney. Using population-based search and evolutionary game theory to acquire better-response strategies for the double-auction market. In *Proceedings of IJCAI-05 Workshop on Trading Agent Design and Analysis (TADA-05)*, 2005.
32. S. Phelps, M. Marcinkiewicz, S. Parsons, and P. McBurney. A novel method for automatic strategy acquisition in n-player non-zero-sum games. In *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'06)*, pages 705–712, New York, NY, USA, 2006. ACM Press.
33. S. Phelps, P. McBurney, S. Parsons, and E. Sklar. Co-evolutionary auction mechanism design: a preliminary report. In *Proceedings of the Workshop on Agent Mediated Electronic Commerce IV (AMEC IV)*, 2002.
34. S. Phelps, S. Parsons, and P. McBurney. An evolutionary game-theoretic comparison of two double-auction market designs. In *Proceedings of the Workshop on Agent Mediated Electronic Commerce VI (AMEC VI)*, 2004.
35. S. Phelps, S. Parsons, E. Sklar, and P. McBurney. Using genetic programming to optimise pricing rules for a double auction market. In *Proceedings of the Workshop on Agents for Electronic Commerce,*, Pittsburgh, PA, 2003.
36. C. R. Plott and V. L. Smith. An experimental examination of two exchange institutions. *The Review of Economic Studies*, 45(1):133–153, February 1978.
37. J. Rust, J. H. Miller, and R. G. Palmer. Behaviour of trading automata in a computerized double auction market. In Friedman and Rust [10], chapter 6, pages 155–199.
38. R. A. Schwartz, J. A. Byrne, and A. Colaninno, editors. *The New NASDAQ Marketplace*. Zicklin School of Business Financial Markets Series. Springer, New York, 2007.
39. A. Shah. Competing exchanges: The international dimension. `http://www.mayin.org/ajayshah/MEDIA/1997/cmarkets.html`, 1997.
40. V. L. Smith. An experimental study of competitive market behaviour. *Journal of Political Economy*, 70(2):111–137, April 1962.
41. E. Sodomka, J. Collins, and M. L. Gini. Efficient statistical methods for evaluating trading agent performance. In *Proceedings of the 27th Conference on Artificial Intelligence*, pages 770–775, Vancouver, Canada, July 2007.

42. P. Stone and A. Greenwald. The first international trading agent competition: Autonomous bidding agents. *Electronic Commerce Research*, 5(2):229–265, 2005.
43. R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA, 1998.
44. `http://www.sics.se/tac`.
45. `http://www.marketbasedcontrol.com/blog/`. Entry for July 26th 2007.
46. P. Vytelingum, I. A. Vetsikas, B. Shi, and N. R. Jennings. IAMwildCAT: The winning strategy for the TAC Market Design Competition. In *Proceedings of 18th European Conference on Artificial Intelligence*, pages 428–434, Patras, Greece, 2008.
47. W. Walsh, R. Das, G. Tesauro, and J. O. Kephart. Analyzing complex strategic interactions in multi-agent systems. In P. Gmytrasiewicz and S. Parsons, editors, *Proceedings of 2002 Workshop on Game-Theoretic and Decision-Theoretic Agents (GTDT-02)*, Edmonton, Alberta Canada, July 2002. AAAI.
48. W. E. Walsh, D. C. Parkes, and R. Das. Choosing samples to compute heuristic-strategy Nash equilibrium. In *AAMAS 2003 Workshop on Agent Mediated Electronic Commerce*, Melbourne, Australia, 2003.
49. M. P. Wellman, S.-F. Cheng, D. M. Reeves, and K. M. Lochner. Trading agents competing: Performance, progress, and market effectiveness. *IEEE Intelligent Systems*, 18(6):48–53, 2003.
50. B. Widrow and M. E. Hoff. Adaptive switching circuits. In *1960 IRE Western Electric Show and Convention Record*, pages 96–104. August 1960.
51. P. R. Wurman, W. E. Walsh, and M. P. Wellman. Flexible double auctions for electronic commerce: Theory and implementation. *Decision Support Systems*, 24(1):17–27, November 1998.
52. P. R. Wurman, M. P. Wellman, and W. E. Walsh. A parametrization of the auction design space. *Games and Economic Behavior*, 35:304–338, 2001.
53. W. Zhan and D. Friedman. Markups in double auction markets. *Journal of Economic Dynamics and Control*, 31(9):2984–3005, September 2007.

A MetroCat: A simple, but powerful, design for CAT games

We developed `MetroCat`, a market mechanism that instantiates the parameterized framework in Section 3, based on several insights about the CAT game. In particular:

- It is crucial to maintain a high transaction success rate, since this rate is not immediately affected by the performance of other markets in contrast to market share and profit share. Thus a strong shout accepting policy, which only allows those shouts that are likely to match with other shouts, is desirable.
- Registration and information fees should be avoided, for these fees cause losses to extra-marginal traders and drive them away. Keeping extra-marginal traders in the market allows them to contribute through their impact on market share.
- Moderate charges on shouts, transactions, and trader profit only impact intra-marginal traders, and because of this they still stay with the market as long as they can make a considerable amount of profit through transactions after covering fees.

These insights led us to develop a CDA-based market mechanism, which uses a history-based shout accepting policy, denoted as AH. AH is based on the GD trading strategy [13]. GD selects a price that maximizes the expected payoff, assuming that, for a given ask price a,

- if another ask price $a' < a$ was offered and was not accepted by a seller, a would not be accepted either;
- if another ask price $a' > a$ was offered and accepted by a seller, a would have been accepted as well; and
- if a bid price $b > a$ was offered in the market, a would have been accepted.

Based on these assumptions, the probability of a being matched is calculated as:

$$Pr(a) = \frac{\sum_{d \geq a} MA(d) + \sum_{d \geq a} B(d)}{\sum_{d \geq a} MA(d) + \sum_{d \geq a} B(d) + \sum_{d \leq a} RA(d)}$$

where

- $MA(d)$ is the number of asks with price d that have been matched;
- $RA(d)$ is the number of asks with price d that were not matched; and
- $B(d)$ is the number of bids with price d.

It is not realistic to keep a full history of shouts and transactions, so GD maintains a sliding window and only considers those shouts and transactions in the window. Computed like this, $Pr(a)$ is a monotonic decreasing function, since the higher a is, the lower $Pr(a)$. It is also assumed that when $a = 0.0$, $Pr(a) = 1$, and there is a certain value u_a, when $a > u_a$, $Pr(a) = 0$. The probability $Pr(b)$ of a given bid being accepted is computed analogously.

AH uses exactly $Pr(a)$ and $Pr(b)$ to estimate how likely a shout would be matched, and only accepts those shouts with a probability higher than a specified threshold $\lambda \in [0,1]$. When it is close to 1, the restriction may become too tight for intra-marginal traders to be able to place shouts in the market. When it is close to 0, the restriction may become so loose that extra-marginal traders are able to place shouts that do not stand much chance of being matched. The former would cause both the market and the traders to lose part of the expected profit and lead those traders to leave, and the latter would cause a low transaction success rate. `MetroCat` uses $\lambda = 0.5$, which we found to be optimal for a game configuration similar to CAT 2007.

In addition to AH, `MetroCat` uses a simple charging policy that imposes low, fixed fees on shouts, transactions, and trader profit, and no charges on registration and information. The fees `MetroCat` imposes on shouts, transactions, and trader profit are respectively 0.1, 0.1, and 10% during the post-tournament experiments described in Section 6.

As described above, `MetroCat` was supplied to all entrants to the 2007 Market Design game as an example specialist that was part of the JCAT package. Given the performance of `MetroCat` in our experiments, it seems that had any entrant used it, even without modification, they would have won.