

An overview and evaluation of the CAT Market Design competition

Tim Miller

Department of Computing and Information Systems

University of Melbourne, Parkville Victoria 3010 Australia

tmiller@unimelb.edu.au

Jinzhong Niu

Stella and Charles Guttman Community College, CUNY

50 West 40th Street, New York, NY 10018, USA

jniu@gradcenter.cuny.edu

Martin Chapman

Department of Informatics, King's College London, Strand WC2R 2LS UK

martin.chapman@kcl.ac.uk

Peter McBurney

Department of Informatics, King's College London, Strand WC2R 2LS UK

peter.mcburney@kcl.ac.uk

An overview and evaluation of the CAT Market Design competition

ABSTRACT

The rise of online commerce has led to a new, emerging discipline at the intersection of economics and computer science, a discipline which studies the properties and dynamics of automated trading in online marketplaces. The CAT Market Design Tournament was created to promote research into the design and deployment of economic mechanisms for such online marketplaces, particularly mechanisms able to adapt automatically to dynamic competitive environments. This research competition has now run for five years (2007–2011), and over that time, been won by four different teams, and had entrants from 13 different countries. In this paper, we describe the motivation and history of the tournament, and present research that has arisen from the tournament. We experimentally “play-off” the five winning entrants of the tournament to evaluate whether the state-of-the-art in automated mechanism design has improved over the course of the CAT competition. The results of our evaluation show a clear and consistent improvement in the performance of CAT winners over the last five years, supporting the belief that the competition has encouraged research in the field.

1 INTRODUCTION

Since 2007, a team of academics from the University of Liverpool, City University of New York, and the University of Melbourne, have run the annual CAT Market Design Tournament, or CAT

Competition. This competition, which is part of the larger Trading Agents Competition (TAC), aims to encourage research into the design and application of automated market mechanisms that are able to adapt automatically to changing competitive environments.

In computer science, there are many research areas where experimental platforms have been developed for comparing and bench-marking research, and open competitions have been built around some of these platforms. Examples include the MiniZinc competition for finite domain constraint solvers¹, and the various TAC competitions for automated trading, supply-chain networks, online ad exchanges, and brokerage of electric power.² From the TAC Competitions, the research arising on automated trading, price prediction and adaptive bidding has been summarised in the book by Wellman et al. (2007). Perhaps the most famous series of research competitions in computer science is *Robocup*, which since 1997 has organised annual competitions for soccer-playing robots, in separate leagues comprising physical robots of various shapes and sizes, and virtual players and managers.³ This crowd-sourced competition series has to date generated no fewer than 15 annual volumes of refereed research papers, each published in Springer's *Lecture Notes in Computer Science* series, as well as numerous other publications, and advanced computer science considerably. The TREC conference on information retrieval⁴ is another example of a competition and online repository, in this case aimed at encouraging and comparing research into information retrieval. This annual competition, started in the early 1990s, aims to promote research into a range of information retrieval tasks, and provides sets of data for researchers to use. However, work from Armstrong et al. (2009) casts doubt on the success of the TREC tournaments. Their research shows that the baselines used in information retrieval experiments are typically uncompetitive — in many cases they perform worse than the first TREC experimental systems —, and there is no upward trend in scores, with results reported in 2008 showing no improvement over results from 1999, and many 2008 scores falling below the median of the original TREC experimental systems. Unlike the CAT competition, entrants are not requested to upload their entries.

The aim of this chapter is to present our experience with the CAT competition (Sections 2), to present details of research that has arisen from the tournaments (Section 3), and to evaluate whether

the CAT competition has helped to improve the state-of-the-art in automated mechanism design (Section 4), so that research in this area does not fall the way of information retrieval. To evaluate the CAT competition, we experimentally compared the five annual winning entrants of the competition (2007-2011) against each other in a series of five-player experiments, as well as a series of two-player round-robin experiments. The results indicate that later entrants perform better on aggregate, especially in the five-player experiments. In the two-player round-robin experiments, the most notable improvement came from the winner in the the 2nd year of the competition, which is the first year that entrants had access to a previous years' entrants to run their own experiments.

2 THE CAT TOURNAMENT

2.1 Motivation

The economic world we inhabit has an enormous diversity of marketplaces. From garage sales of unwanted household goods to local *bodegas* and corner stores, to supermarkets and international chains, through automated online exchanges for the trading of complex derivative and futures products, this diversity is astonishing. The response of mainstream economic theory to this diversity until recently, however, was to abstract all the particular features of markets away to create a generic model, a model of a marketplace having independent buyers and sellers who interact in a uniform manner (which is often left implicit). This generic model ignores many of the specific features of particular markets: for example, that in markets for medical services, the provider and the purchase decision-maker may often be the same person (e.g., a doctor acting on behalf of a patient), or that in higher-education markets, the consumer — the student — may also constitute the product being purchased, or that in management consultancy markets, there may be no way for the purchaser to assess the quality of the provided service, neither ahead of time nor even afterwards. Most importantly, a generic model also abstracts away from the “ecology” of all the different markets operating in broad parallel with one another, some competing directly with others, some competing indirectly, some parasitic on others, some complementary to others, and some

generating information used by others (Mirowski 2002). Ignoring this ecology of markets means that, for example, with a single generic model of a marketplace we cannot explain why the world's most sophisticated futures trading exchanges in Chicago or New York or London operate within only short walking distances from local household garage sales, and yet use very different rules and procedures. Of course, such complex futures markets usually target different trading participants to household garage sales, but this is precisely why the rules of the two types of markets are different. A generic model of all markets does not allow for different markets to have different rules, nor does it seek to understand the impacts and dynamics of these rules, and how they may interact with one another.

Although mainstream economics has long studied some exceptions to the generic model of perfect competition — for example, monopoly, oligopoly, and monopsony markets — the discipline has mostly ignored competition *between* marketplaces. But competition between markets has grown in importance with the rise of the Internet, the Web, and electronic trading. With both buyers and sellers now able to be located anywhere, so too may their transactions take place anywhere. As a result, the precise rules and procedures of a particular market may have important consequences for the attraction and retention of buyers or sellers, with each trader able to choose between multiple competing marketplaces. The strong opposition demonstrated recently by the City of London and the British Government to the so-called *Tobin Tax* on financial transactions, for instance, is based on a justified belief that any such tax, implemented locally and not globally, will lead to a loss of business from those exchanges that are taxed to those which are not.

One key motivation, therefore, for establishing the CAT Market Design Tournament was to better understand the nature of competition between marketplaces, in the absence of any mature economic theories of such competition. In particular, such an understanding should provide scientific understanding and guidance on how specific market rules and procedures would impact on market and trader performance, and how they would interact with or influence one another via competing markets. It is likely that the specific network topology of connections between markets operating in the same ecology would also impact on marketplace and trader actions and performance. The

operators of, and traders in, the world's stock exchanges have lots of anecdotal understanding of these relationships and impacts, but no coherent, grounded and granular theory exists for the specific rules that exchange operators typically consider and deploy. Such a theory would provide guidance to the designers, implementers and managers of electronic marketplaces, people who are increasingly information technologists rather than economists.

Secondly, given the rise of electronic trading and the globalisation of trading, it is easy to imagine that operators of exchanges would need to respond in real-time, or close to real-time, to changes in their business environments: If a competing exchange introduces a new rule or new fee structure, and quickly attracts more traders or traders who are more desirable, then rival exchanges losing custom will need to respond as quickly as possible. Thus, our second key motivation was to develop the science and software engineering of adaptive market mechanism design, aiming ultimately for fully automated mechanism design.

Traditionally, the design of market mechanisms has been undertaken by human economists using a complex mix of mathematical analysis (chiefly using game theory), computer simulation, experiments with human subjects, and trial operations. Edward Nik-Khah, for example, observes that all of these methods were used in the design of the specific rules of the US Federal Communications Commission's auctions of PCS radio frequency spectrum in the mid 1990s (Nik-Khah 2005). Given the subtlety and complexity of actual market design engineering, it was not obvious how, or even whether, automated market mechanism design would be possible. We developed the CAT Market Design Tournament to encourage multi-disciplinary research in this area, and to provide a standard experimental platform by which such research could be compared and bench-marked. In terms of the refereed research publications generated by the game participants, our goal of encouraging research here has certainly been achieved, and several promising approaches to adaptive and automated mechanism design proposed, for example, Niu et al. (2012), Pardoe et al. (2010), and Phelps et al. (2010).

In designing the CAT Tournament and its supporting software platform *JCAT*, we sought to model practice in actual (spot) financial exchanges, while keeping the game rules straightforward. Thus,

the marketplaces in CAT are called *specialists*, following the standard stock market terminology. Similarly, the performance criteria used in the Tournament reflect those used by managers to evaluate performance of real-world providers of exchange services: profits, market shares, and transaction success rates (i.e., matching intending traders with appropriate counter-parties). The strategies used by software traders in the CAT platform are four automated trading strategies commonly studied in the trading agents literature, and used, to a greater or lesser extent, by automated traders in financial markets. Further details are presented in the next sub-section.

Full details of the organisation of the CAT Market Design Tournament can be found on the CAT web-page⁵ and in the technical documents describing the game and participation (Cai et al. 2009; Niu et al. 2009). In addition, the various research publications published by those involved in the CAT Tournament provide useful information about the game operations as well as reports of the resulting research; the most important of these research publications are discussed in Section 3. The next two sub-sections describe the operation of the Tournament, and its history, respectively.

2.2 Operation of the Tournament

We now briefly outline the structure of the CAT Tournament. Entrants to the competition design and implement software agents, called *specialists*, which operate as matchmakers between automated traders. Specialists are competing marketplaces, in the same way that, say, the New York Stock Exchange, NASDAQ, the London Stock Exchange, the Paris Bourse, etc are all competing markets. The traders are adaptive software agents provided by the Tournament organisers, and which buy and sell a single generic commodity. Traders are either buyers or they are sellers but not both. Each trading day, each trader first selects a specialist to trade through on that day, and then decides what level of shout to bid or ask in that specialist. For the shout decision, each trader is endowed with one of four standard automated trading strategies:

- ZI-C: Zero Intelligence–Constrained (Gode and Sunder (1993)), a strategy which essentially involves random shouting (bidding by potential buyers or asking by potential sellers), within

the budget constraint of each trader. This strategy ignores any past trading history or experience of the trader.

- ZIP: Zero Intelligence Plus (Cliff and Bruten (1997)), a strategy which takes account of the trader's past experience to extend ZI-C, and adapts the shouting strategy according to changing market conditions.
- RE: Roth and Erev (Erev and Roth (1998)), a strategy which mimics reinforcement learning by humans in game-playing.
- GD: Gjerstad-Dickhaut (Gjerstad and Dickhaut (1998)), a strategy in which traders form beliefs about the expected success and profits from particular shouts based on their past trading experience.

These strategies were selected because they have been extensively studied in the automated trading literature and some have also been shown to work well in practice. Within the CAT Tournament, the strategy used by each trader is private knowledge, and is not revealed to other traders and nor to the specialists. Also, the proportion of traders executing each of the four strategies is not revealed to specialists.

Each trading day in the CAT Tournament begins with specialists deciding on and posting their fee levels for all traders to see. The fee structure and fee levels inform the traders' decisions as to which specialist to select that day. Specialists have the following goals:

- To attract traders to their marketplace, in competition with the other specialists;
- To match potential buyers with potential sellers through a market matching and clearing algorithm; and
- To earn profits by deciding a daily fee structure and fee levels, such as fees for placing a shout, or a commission fee based on a trader's profit.

Entrants to the Tournament are scored on three equally-weighted criteria each trading day: share of all profits earned by specialists that day; share of traders attracted to their marketplace that day; and the specialist's transaction success rate, which is the percentage of traders that were successfully matched with counter-parties by the specialist. These daily scores are summed over the Tournament, with the winner being that specialist having the highest cumulative score across all scored days. To eliminate the possibility of specialists taking tactical advantage of edge effects, scoring in the Tournament does not commence until some days after the first trading day, and ends some days before the last trading day.

2.3 History of the Tournament: 2007–2011

The CAT Market Design Tournament has now run five times, once each year during 2007–2011. To date the CAT Tournament has been operated as part of the annual Trading Agent Competitions (TAC), hosted each northern summer by the International Trading Agent Research Association.⁶ The first TAC game, now called TAC Classic, invited entrants to program trading agents able to buy and sell various electronic goods in multiple, complementary marketplaces. The name *CAT* was adopted for the Market Design Competition for two reasons: firstly, because it is short for *catallactics*, the science of economic exchange, and secondly, because the game operates in reverse to the way in which TAC Classic operates: in CAT, the entrants provide marketplaces and the competition organisers provide automated traders, rather than the reverse.

The TAC Competitions are usually held alongside a leading computer science research conference in artificial intelligence, multi-agent systems, or e-commerce, and so move location each year. The CAT Tournament has been hosted in Vancouver, Canada (2007), Chicago, IL, USA (2008), Pasadena, CA, USA (2009), Boston, MA, USA (2010) and Barcelona, Catalonia, Spain (2011). The games in 2007, 2008 and 2009 were operated from a server in the Department of Computer Science at the University of Liverpool, UK, while those in 2010 and 2011 were operated from a machine in what is now the Department of Computing and Information Systems, at the University of Melbourne, Australia.⁷

Entrants to the CAT Tournament have come from around the world, and have included both large academic research team and individuals. It is a measure of the intellectual challenges provided by the game that individual entrants from developing countries having clever ideas may compete effectively against first-world teams having access to the large-scale computational processing power needed for full testing of entrants. Table 1 shows the number and nationalities of entrants in each year of the CAT Tournament.

Year	Entries	Winner (Country)	Countries Entering
2007	10	IAMWildCat (UK)	Australia, Croatia, France, Greece, Iran, Romania, UK, USA
2008	11	PersianCat (Iran)	Australia, Croatia, Egypt, France, Greece, Iran, UK, USA
2009	14	jackaroo (Australia-France)	Australia, China, Croatia, France, Greece, Iran, Mexico, Taiwan, UK, USA
2010	9	Mertacor (Greece)	Australia, Croatia, Egypt, France, Greece, Iran, Poland, Taiwan, UK, USA
2011	8	jackaroo (Australia-France)	Australia, France, Greece, Mexico, Poland, Taiwan, UK

Table 1: Summary of CAT Market Design Tournaments 2007–2011

As with the other TAC games, entrants to CAT each year are encouraged to upload their entries (either as raw code or compiled binaries) to an online repository following the tournament.⁸ Doing this enables future entrants to test their intended entries against those of past entrants, and helps ensure that the research community as a whole learns from the crowd-sourced research experiment that the CAT Tournament effectively constitutes. We have drawn upon the entries available in this repository to run the Hall-of-Fame simulation experiments we report in Section 4.

3 RESEARCH ARISING FROM THE TOURNAMENT

Much research work has arisen from the design and the organisation of the CAT Tournaments, exploring issues such as the modelling and engineering of electronic market mechanisms, the design of novel market rules, the systematic evaluation of their performance, the dynamics of the interactions between multiple marketplaces, and between markets and traders, and what can be learnt from these dynamics to adapt market mechanisms to achieve better performance.

In this section, we summarise some of the most important research that has resulted from the CAT Tournament. We broadly divide this into two categories: research on market mechanisms, and research on automated traders using market mechanisms. In addition to the publications discussed below, no fewer than four doctorates have now been awarded for research arising from the CAT Tournament or which uses the JCAT platform: Furuhata (2009), Niu (2011), Robinson (2011), and Shi (2011).

3.1 Modelling market mechanisms

Niu et al. (2008) presented a modular model of market mechanisms in the effort of preparing a reference implementation of market mechanism for entrants of the CAT Tournaments. In this modular view, which extends prior understandings of market mechanisms, e.g., (Wurman et al. 1998), a market mechanism is composed of multiple auction rules, or what they called *policies*, each regulating one aspect of the marketplace. There are seven different types of policy:

1. *matching policies* define how a market matches shouts made by traders;
2. *quote policies* determine the quotes issued by markets;
3. *shout accepting policies* judge whether a shout made by a trader should be permitted in the market;
4. *clearing policies* define when to clear the market and execute transactions between matched asks and bids;

5. *pricing policies* set transaction prices for matched ask-bid pairs;
6. *charging policies* determine the charges imposed by a market; and
7. *subscribing policies* control the collection of market information so that the market mechanism may adapt itself.

A more detailed description of this model and numerous policies that can be classified in such a way and are implemented in JCAT is given by Niu (2011). According to the entries that were uploaded into the TAC repository, this modular model has been generally adopted by the entrants of the CAT tournaments and the policies in JCAT have been the primitive building blocks for the entrants to design more advanced components of their market mechanisms.

For example, a common shout accepting policy is what in Niu (2011) calls a *quote-beating accepting policy* (also known as the NYSE shout improvement rule due to its use by the New York Stock Exchange). This policy is typically used in the *continuous double auction* (CDA), where traders may submit shouts any time during trading, transactions are executed whenever a pair of shouts can be matched, and a shout is allowed to be placed in the market only when it is more competitive than the corresponding market quotes — ie, proposed asks need to be lower than the ask quote, and proposed bids need to be higher than the bid quote. The market quotes are typically determined by the least-competitive matched shout from the opposite side and the most-competitive unmatched shout from the same side, i.e., for example, the ask quote being based on the lowest matched bid and the highest unmatched ask (McCabe et al. 1993). The quote-beating policy is believed effective in speeding up the convergence to the market competitive equilibrium in a CDA market (Cliff and Bruten 1997). Although this policy is simple and can be easily understood by human traders in traditional markets, it relies upon limited information and may lead to volatile transaction prices when traders are sub-rational and the shouts they place do not cluster around the market competitive equilibrium price. In fact, the market quotes can be viewed as estimates of the market competitive equilibrium price, the ask quote being the upper bound and the bid quote being the lower bound.

Niu et al. (2006) proposed what they called the *equilibrium-beating accepting policy* using the average of transaction prices in a sliding window of the transaction history as the estimated equilibrium price. They showed that this simple method was effective in reducing price fluctuations as well as improving the allocative efficiency in a stand-alone CDA market populated by ZI-C traders. Various CAT entries devised their variants to obtain desirable results in the CAT setting. PersianCat in CAT 2008 (Honari et al. 2008), for instance, estimated the market equilibrium price with a running average of the minimum price of matched bids and the maximum price of matched asks on each day over a certain number of recent game days. This variant of equilibrium estimation samples from a relatively long history of transactions and helps to maintain an even smoother estimate over time. An improved version of AstonCAT in CAT 2010 (Chang et al. 2011) combined the original equilibrium-beating accepting policy and the PersianCat variant in a linear manner, where the weighted parameters are carefully chosen to balance the information learnt from the intraday trading as well as trading spanning across multiple days. PersianCat won the 2008 CAT tournament and the improved AstonCAT, or AstonCAT-plus, performed well in controlled experiments, taking a close second place in 2010 after Mertacor, the CAT 2010 champion, and beating all other entries by a large margin. The designers of both PersianCat 2008 and AstonCAT-plus attributed their specialist's good performances to their effective market equilibrium estimation algorithms, which were the core component of many policies in their respective market mechanisms rather than merely the accepting policies. For further details on the design of these entries and some others, see the references in Table 2.

The complex interactions between multiple marketplaces and various types of traders and the limited coverage during actual tournaments make it necessary to run post-tournament analyses to understand the multi-marketplace dynamics, identify causal effects, and provide guidance for further improvements. Niu et al. (2010) systematically carried out two types of analyses, *white-box analysis* and *black-box analysis*, of the CAT 2007 entries. The white-box analysis ran experiments in the same way as the actual tournament ran and tried to relate observed market dynamics to

Year	Entry	Publications
2007	CrocodileAgent	Petric et al. (2008)
	IAMwildCAT	Vytelingum et al. (2008)
	PersianCat	Honari et al. (2008)
2008	Mertacor	Stavrogiannis and Mitkas (2009)
	PersianCat	Honari et al. (2008)
	PSUCAT	Sohn et al. (2009b)
2009	jackaroo	Furuhata et al. (2009)
	Mertacor	Stavrogiannis and Mitkas (2010)
2010	AstonCAT	Chang et al. (2011)

Table 2: Publications on the design of entries of the CAT tournaments

the policies adopted by those entries, particularly how intra-marginal traders — those traders that would transact if all traders place truth-telling shouts and the market is cleared at the equilibrium — and extra-marginal traders — all non-intra-marginal traders — migrate between marketplaces as the entries interact and adapt their strategies over time. The black-box analysis considered the behaviours of entries as atomic actions in the game-theoretic sense and played these actions repeatedly across many games in an ecological process where highly-rewarded actions become more and more popular and lesser-rewarded actions die out gradually, in a way similar to the simulated ecological evolution of entries in the Iterated Prisoner’s Dilemma Tournament of Axelrod (2006). This ecological process helped to identify weaknesses of certain entries in varying competitive environments. For example, IAMwildCAT, the CAT 2007 champion, may lose the competition against PersianCat through one-on-one games in certain experimental setups. Indeed the actual tournaments reflected only limited trajectories of interaction dynamics and were not able to tell the full story.

Robinson et al. (2010), in particular, examined whether the performance of entries from CAT 2008 and 2009 could be *generalised* when the competitive environment varied, such as when facing different market competitors, or when the distribution of strategies varied across the trader population. They showed that the rankings of entries was sensitive to changes in these competitive settings. This result was further confirmed by results obtained in other post-tournament analyses (Vytelingum et al. 2008; Stavrogiannis and Mitkas 2009; Stavrogiannis and Mitkas 2010; Chang

et al. 2011). Chen and Yu (2012) have also studied the effect of market policies on market performance indicators, using the JCAT platform.

Such results are not surprising as CAT games involve complex interactions between adaptive marketplaces and traders, and may not allow solution concepts in game theory like dominant strategies or Nash equilibrium strategies. Nevertheless, empirical analysis still has merit in examining the relative robustness of entries facing varying situations and helps to inspire and foster further improvements.

As stated earlier, one of the ultimate goals of running the CAT tournaments is to promote research on automated auction mechanism design. Here we describe two pieces of work that exemplify early steps towards achieving this goal. Gruman and Narayana (2008) devised a method, based on machine learning, for estimating the distribution of different trading strategies among traders in CAT Tournaments. The specific classes of trading strategies used by CAT traders are known to CAT entrants, but not the strategy used by any particular trader, nor the distribution of the different strategy types across the trader population. Entrants may be able to optimise their policies if the distribution can be accurately estimated. In this work, the authors showed that a classifier based on hidden Markov models was able to estimate the distribution of trading strategies with an accuracy of around 60%, better than other classifiers they evaluated (including support vector machines and conditional random fields), and in contrast to the 25% accuracy expected when guessing randomly. From this result they designed a framework based on Markov decision processes for entries in CAT Tournaments. Although this framework turned out to be not as successful, this effort helped to shed light on the difficulty of modelling the multi-marketplace scenario and encouraged further exploration in this direction.

Another piece of work along this avenue is the grey-box approach of Niu et al. (2012), based upon their prior work using a white-box analysis and a black-box analysis. Combining the two analysis methods, this grey-box approach tackled the problem of automated mechanism design through automated search in a tree-based structure of possible auction mechanisms. The grey-box approach used reinforcement learning to explore the composition of the structure, related the performance of

auction mechanisms to that of auction rules that formed the mechanisms, and utilised evolutionary computation to maintain viable auction mechanisms. Experiments showed that the grey-box approach was able to produce significantly better performance than a classic genetic algorithm. This approach is particularly useful due to the fact that CAT entries typically adopt a modular design framework, as this not only makes it easier to run post-tournament empirical analysis and evaluate the performance of entries on a common basis, but also make it possible to combine policies from different entries and to generate and search for better variants of these entries.

Finally, the CAT Tournament environment provides a suitable setting to study economies involving competing or networked marketplaces (Cai et al. 2008; Cai et al. 2010; Moyaux et al. 2010). These many publications show the value of the CAT Tournament and the supporting software platform, JCAT, in defining a common set of research problems and a shared basis for exploring them.

3.2 Modelling traders in multi-marketplace environments

Faced with multiple competing marketplaces, a trader faces two decision problems: which market to select and what amount to bid or ask in the selected market. For self-interested traders, both decisions can be undertaken with the intention of maximising expected profits. These two questions have led to two lines of research.

3.2.1 Marketplace selection

The CAT game designers modelled each trader as an N-armed bandit problem solver, where the daily profits of a trader were used as rewards for the trader to learn to optimise their selection between alternative action-options (ie, marketplaces) (Niu et al. 2007). They examined the model through experiments running marketplaces that charge differently and showed that profitable traders consistently favoured that marketplace charging the lowest fees. This configuration of market selection has been used in all CAT Tournaments; similar results were observed in the post-tournament analysis (Niu et al. 2008), where the movement of intra-marginal and extra-marginal traders between CAT 2007 entries were examined.

Sohn et al. (2009a) created a simple game-theoretic model and explored the Nash equilibrium strategies for a few pairs of intra-marginal and extra-marginal truth-telling traders choosing between competing marketplaces. Their results demonstrated that extra-marginal traders are more sensitive than intra-marginal traders to registration fees, fees that traders need to pay when they choose a marketplace to trade in no matter whether they are able to transact or not. In later work, Sohn et al. (2010) obtained similar results when software trading agents were replaced by human traders. Shi et al. (2012) further investigated the problem of market selection, presenting a formal description of the problem and Bayes-Nash Equilibrium (BNE) solutions. In particular, after certain assumptions making the problem analytically tractable, they showed that there existed BNEs where all traders choose the more expensive marketplace out of two, or where traders choose that more expensive marketplace stochastically with a non-zero probability. Further analysis based on evolutionary game theory revealed that the latter BNE was neither stable, nor easy to reach. On the other hand, reaching the former BNE depended on particular pairings of sellers with buyers, and these pairings depended on chance factors such as the numbers of traders, their initial distribution, and the fees charged by the marketplaces.

Miller and Niu (2012) explored the parameter spaces of several N-armed bandit algorithms for the purpose of market selection. They ran two sets of experiments, one containing only adaptive specialists and one containing only non-adaptive specialists, and found that the softmax strategy (Sutton and Barto 1998) is the most successful at maximising trader profit and overall market efficiency in both adaptive and non-adaptive markets.

3.2.2 Trading strategies

Although most research on trading agents in the CAT Tournaments has focused on market selection, there is some work that paid attention to trading strategies: strategies that traders are equipped with for making offers in marketplaces. Sohn et al. (2009b) argued that trading strategies for traders in CAT tournaments, such as ZIP, originally designed to perform in a single, stand-alone marketplace using some presumed market mechanism, may not work well when they need to ac-

complete additional decision making on market selection, and when specialists entries typically adopt varying customised market mechanisms. They discussed the misalignment of ZIP with the pricing policy of PSUCAT, their entry in CAT 2007, and showed that a revised version of ZIP could achieve higher profits and better choose marketplaces.

Scope exists for considerably more research in this area, particularly looking at trading strategies other than the four used in the Tournaments, and in developing strategies that adapt to or exploit the particular mechanisms adopted by the selected specialist.

4 HALL-OF-FAME SIMULATIONS

In this section, we present a lightweight evaluation of the CAT tournament. The hypothesis of CAT is that running competitive tournaments helps to improve the state-of-the-art in automated mechanism design. It is not possible to directly test this hypothesis, because other factors will also influence research into this topic. Instead, we indirectly test this hypothesis by measuring the relative success of participants in the CAT tournament over the past five years. These experiments are not intended to be conclusive or complete, but instead to give us an indication of the effectiveness of the CAT competition.

We present two sets of experiments in this section:

1. Hall-of-fame grand tournament: the first set of experiments consists of 20 CAT games, in which the entrants are the five winners of the CAT tournament over the period 2007–2011 (see Table 1).
2. Hall-of-fame round-robin: the second set of experiments takes the same five entrants, but plays each of them off against the other 20 times in a round-robin tournament of two-player CAT games; that is, we have 10 different pairs, each playing 20 games against each other.

By measuring various outcomes over the course of these two sets of experiments, we would expect to see a positive correlation between specialist performance and the year in which they won the

tournament; that is, winners from later years should perform better on aggregate than those from earlier years.

4.1 Experiment Design

We use the JCAT platform version 0.17⁹ to run CAT games as part of the experiments. Our experiments were setup to replicate a CAT Tournament game, meaning that each game consisted of 500 trading days, with 10 rounds per day, and 400 traders, evenly split between buyers and sellers. The distribution of the four trader types described in Section 2 is also uniform (25% each). Private values of the generic commodity being traded are randomly drawn from a uniform distribution between 50 and 150. These parameters are varied during CAT tournaments, with the entrants not knowing their values or even their distributions, but for the experiments, we maintain the same values.

To measure the performance of each specialist, we use the same three measures used to score the CAT tournament (Cai et al. 2009). On each trading day, the following three measures are taken for each specialist:

1. **Profit share:** the profit obtained by a specialist as a proportion of the total profit made by all specialists in the market, between 0 and 1 inclusive.
2. **Market share:** the market share obtained by a specialist as a proportion of all traders that registered with a specialist on that day, between 0 and 1 inclusive.
3. **Transaction success rate:** the number of shouts that were matched on that day, as a proportion of the total number of shouts placed with that specialist on that day. Note that this is independent of the number of shouts placed with other specialists.

These three measures are added together to produce an overall score for the trading day. The total game score is simply the summation across all trading days.

In addition, we measure the efficiency of individual specialists and the overall market. *Global allocative efficiency* measures how close the entire market is to trading at the equilibrium price,

where the *equilibrium price* is defined as the price at which demand equals supply when all traders offer to buy or sell at their private values, assuming that all traders in the market can trade with each other. The global allocative efficiency is calculated using:

$$E = \frac{\sum_j \sum_i |v_i^j - p_i^j|}{\sum_j \sum_i |v_i^j - p_0|} \quad (1)$$

in which p_0 is the equilibrium price of the market, v_i^j is the private value of trader i in specialist market j , and p_i^j is the price paid by trader i in specialist market j .

Allocative efficiency of individual specialists is similar to the global allocative efficiency defined in Equation 1, but is measured only for transactions in the individual specialist's market.

4.2 Results

4.2.1 Grand tournaments

Figure 1 shows the average game score and average allocative efficiency for each winning entrant (2007-2011), over all 20 grand-tournament experiment runs. Error bars represent the 99% confidence interval. From this figure, there is a clear and consistent upward trend in the scores.

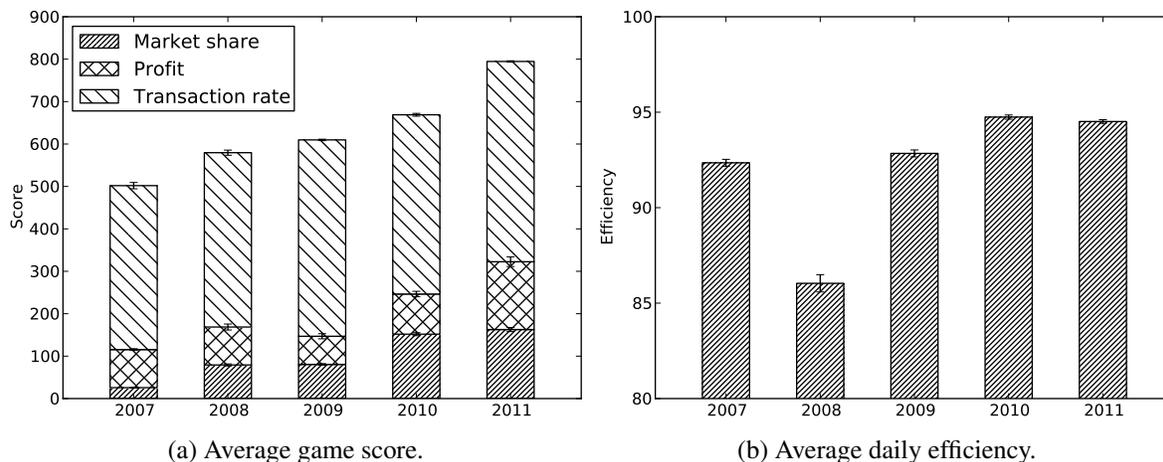


Figure 1: Average score and allocative efficiency for each entrant over all grand tournament experiments.

Table 3 shows the proportion of placings that the specialists received over the 20 experiments that

Specialist	1 st	2 nd	3 rd	4 th	5 th
2007	0	0	0	0	20
2008	0	0	3	17	0
2009	0	1	16	3	0
2010	0	19	1	0	0
2011	20	0	0	0	0

Table 3: Number of placings for each specialist over all 20 grand tournament experiments.

were run. This also demonstrates a clear and consistent trend of more recent specialists outperforming older specialists. The most recent specialist won all 20 games in the experiments, and the oldest came last in all games.

4.2.2 Round-robin tournaments

Figure 2 shows the average game score and average allocative efficiency for each winning entrant (2007–2011), over all round-robin experiment runs. Error bars represent the 99% confidence interval. From this figure, one can see clear upward trend in the data, but that the 2008 and 2009 winners both out-score the 2010 and 2011 winners. All score higher than the 2008 winner. The average efficiency also shows an upward trend, with the notable exception that the 2008 winner trades well below the allocative efficiency of all other specialists. The 2011 entrant achieves an allocative efficiency marginally below that of the 2010 entrant.

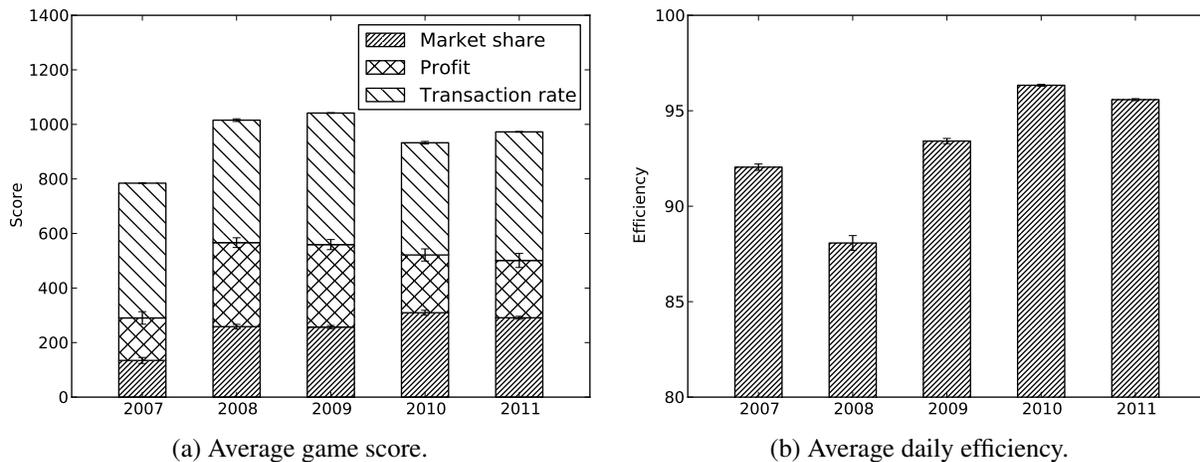


Figure 2: Average score and allocative efficiency for each entrant over all round-robin experiments.

Figure 3 provides a further breakdown of the data from Figure 2, showing the average scores and allocative efficiency for each pair of specialists playing against each other. Two adjacent bars represent a pair, with the left bar representing the “earlier” specialist. The final column in each, the average, is the average of all pairs.

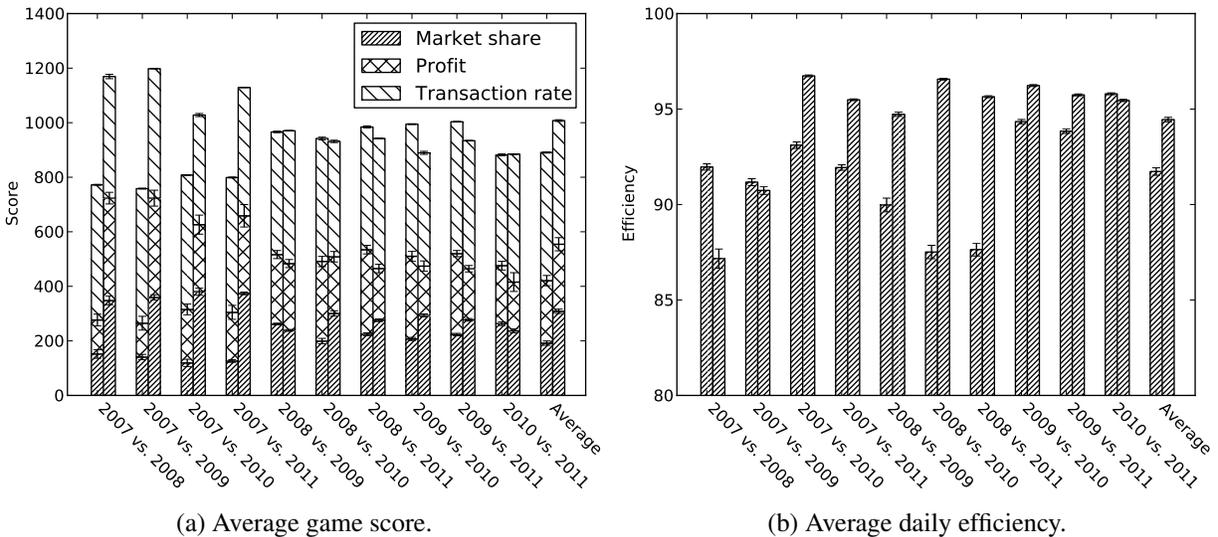


Figure 3: Average total score for specialists in two-player round-robin experiments.

From this figure, we can see that the 2007 winner is substantially out-scored by all other entrants, and that, consistent with Figure 2, the 2009 winner beats all other specialists in the round-robin experiments. Allocative efficiency is consistent with Figure 2, with later specialists outperforming earlier specialists, with the exception of the 2008 winner, which traded below the other specialists in all round-robin experiments. While there is a clear upward trend in the data, the 2008 and 2009 entrants both out-scored the 2010 and 2011 entrants.

4.3 Discussion

These results support our hypothesis that the research in automated mechanism design has improved over the past five years, and given the consistent rise in scores shown in Figure 1 and Table 3, we believe that the CAT tournament is helping to drive this.

The grand tournament experiments demonstrate a clear and consistent upward trend in the scores

of specialists, supporting our hypothesis. The round-robin experiments support the hypothesis, but to less of an extent, with the trend being upward, but some earlier tournament winners outscoring later winners. This is consistent with the findings of (Robinson et al. 2010), who demonstrated that entrants in the 2008 and 2009 competitions are sensitive to the other specialists with whom they compete. It is likely that the specialists we used were tuned to games with several competitors, rather than two-player games.

The graphs in Figures 1–3 show that obtaining a large market share is key to scoring well. Transaction rates are fairly evenly matched over all surveyed specialists. However, the 2008 and 2009 entrants scored well in the round-robin tournaments with increased profits: both had lower market share and similar transaction rates to the 2010 and 2011 specialists, yet out-scored them overall.

4.4 Threats to validity

There are several threats to the validity of the conclusions stated above, which we present in this section.

First, to evaluate whether the CAT competition has fulfilled its aims, we have used only the winning entrants over the life of the competition, ignoring other entrants.

Second, we explore only a slice of the possible parameter space. Such a restriction is necessary because the experiments take weeks to run. Robinson et al. (2010) identify that CAT entrants are sensitive to many factors such as the types of traders and the other specialists with whom they compete. It is therefore possible that, were we to vary the competitive parameters, the results would show less support for our hypothesis.

Third, while we use experiments to assess whether the CAT competition has achieved our aim of encouraging research into automated mechanism design, we cannot measure this directly. Instead, we have measured the entrants in the competition only. Clearly, there are other external factors that influence this research area, and influence the quality of the entrants. Despite this, research arising from this area has made use of specialists in the TAC repository, and we are confident this has helped to improve this research field.

Finally, while the CAT competition aims to encourage research into automated mechanism design in general, the specialists entered into the tournaments are geared towards the CAT Tournaments in particular, and it may be that the CAT competition is not helping achieve progress outside of this particular niche of the Tournament itself. Our opinion is that the research arising from this Tournament is most definitely applicable, but would need to be tuned to particular application domains, in the same way that the CAT specialists have been tuned to the CAT Tournament.

5 CONCLUSIONS

In this paper, we presented an overview the CAT Market Design Tournament, including the history of the tournament and key research that has resulted from the Tournament. The primary aim of the CAT Tournament is to encourage research into automated mechanism design in online marketplaces. We evaluated whether the competition is achieving this aim using a series of experiments, in which the winners of the first five tournaments played-off against each other. In the first set of experiments, the five winners played a series of grand hall-of-fame tournaments against each other, while in the second set of experiments, the winners were pitched against each other in a set of round-robin two-player games.

Results indicate that the design of automated mechanisms for the CAT Tournament has improved consistently over the five years of the competition, both from the perspective of individual markets, and from a global perspective. We conclude that these results support the conclusion that the CAT Tournament is achieving its aims.

We believe that one reason the Tournament has been successful is that, after each competition, entrants upload binary versions of their specialist, allowing other researchers to use these as benchmarks in their own experiments. Such a repository is missing from the TREK repository, which analysis indicates is less successful than the CAT tournament at improving the state-of-the-art (Armstrong et al. 2009).

This chapter has presented a review of the CAT Tournament over the first five years of its exis-

tence, from 2007 to 2011. The CAT Tournament will continue to run in the future, and we are confident that it will continue to support important state-of-the-art research in automated mechanism design, automated trading, and competition between multiple marketplaces. We hope that the CAT Tournament and the other TAC tournaments serve as models for competitions in other research areas.

Acknowledgements

The CAT Tournament was partly funded by the UK EPSRC, through the *Market Based Control of Complex Computational Systems* project (GR/T10657/01 and related grants), and this support is gratefully acknowledged. McBurney is also grateful for financial assistance received from a University of Melbourne Visiting Research Fellowship.

Notes

¹See <http://www.g12.csse.unimelb.edu.au/minizinc/index.html>.

²See <http://www.sics.se/tac/>.

³See <http://www.robocup.org/>.

⁴See <http://trec.nist.gov/>.

⁵See http://www.marketbasedcontrol.com/blog/?page_id=5.

⁶See <http://tradingagents.org/>.

⁷The authors of this paper include the CAT Gamemaster for 2007–2009 (McBurney), the CAT Gamemaster for 2010–2012 (Miller), and the lead software developer of the CAT Platform (Niu). All three of these authors were involved in the design of the tournament, along with others, principally Enrico Gerding, Simon Parsons and Steve Phelps.

⁸See <http://www.sics.se/tac/showagents.php>.

⁹See <http://jcat.sourceforge.net/>.

REFERENCES

- Armstrong, T., A. Moffat, W. Webber, and J. Zobel (2009). Improvements that don't add up: ad-hoc retrieval results since 1998. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pp. 601–610. ACM.
- Axelrod, R. (2006, December). *The Evolution of Cooperation*. New York, NY: Basic Books, Perseus Books Group.
- Cai, K., E. Gerding, P. McBurney, J. Niu, S. Parsons, and S. Phelps (2009). Overview of CAT: A Market Design Competition (version 2.0). Technical Report ULCS-09-005, Department of Computer Science, University of Liverpool, Liverpool, UK. Version 2.0.
- Cai, K., J. Niu, and S. Parsons (2008). On the economic effects of competition between double auction markets. In *Proceedings of the 10th Workshop on Agent-Mediated Electronic Commerce (AMEC 2008)*.
- Cai, K., J. Niu, and S. Parsons (2010). Network effects in double auction markets with automated traders. In *International Workshop on Agent-Mediated E-Commerce (AMEC 2010)*, Toronto, Canada.
- Chang, M., M. He, and X. Luo (2011). AstonCAT-Plus: An efficient specialist for the TAC Market Design Tournament. In *IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, pp. 146–151.
- Chen, D.-N. and J.-Y. Yu (2012). An expansion matching method to improve transaction effectiveness in the double auction market. *E-Commerce Research and Applications* 11(1). (*Forthcoming*).
- Cliff, D. and J. Bruten (1997). Minimal-intelligence agents for bargaining behaviours in market-based environments. Technical Report HPL-97-91, Hewlett-Packard Research Laboratories, Bristol, England.

Erev, I. and A. E. Roth (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* 88(4), 848–881.

Furuhata, M. (2009). *E-Market Mechanism Design for Supply Chain Management*. Ph. D. thesis, University of Western Sydney and University of Toulouse 1, Sydney, Australia and Toulouse, France.

Furuhata, M., L. Perrussel, J.-M. Thévenin, and D. Zhang (2009). Experimental market mechanism design for double auction. In *Proceedings of the 22nd Australasian Joint Conference on Advances in Artificial Intelligence, AI '09*, Berlin, Heidelberg, pp. 1–10. Springer-Verlag.

Gjerstad, S. and J. Dickhaut (1998). Price formation in double auctions. *Games and Economic Behaviour* 22, 1–29.

Gode, D. K. and S. Sunder (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy* 101(1), 119–137.

Gruman, M. L. and M. Narayana (2008). Applications of classifying bidding strategies for the CAT Tournament. In W. Ketter, A. Rogers, N. Sadeh, and W. Walsh (Eds.), *Proceedings of the International Trading Agent Design and Analysis Workshop (TADA 2008)*, Chicago, IL, USA.

Honari, S., M. Gomrokchi, M. Ebadi, A. Fos-hati, and J. Bentahar (2008). Simulating new markets by introducing new accepting policies for the conventional continuous double auction. In *Proceedings of the 2008 Spring simulation multiconference, SpringSim '08*, San Diego, CA, USA, pp. 89–97. Society for Computer Simulation International.

McCabe, K. A., S. J. Rassenti, and V. L. Smith (1993). Designing a uniform price double auction. In D. Friedman and J. Rust (Eds.), *The Double Auction Market: Institutions, Theories and Evidence*, Santa Fe Institute Studies in the Sciences of Complexity, Chapter 11, pp. 307–332. Cambridge, MA: Westview Press, Perseus Books Group.

Miller, T. and J. Niu (2012). An assessment of strategies for choosing between competitive marketplaces. *Electronic Commerce Research and Applications* 11(1). (Forthcoming).

Mirowski, P. (2002). *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge, UK: Cambridge University Press.

Moyaux, T., P. McBurney, and M. Wooldridge (2010). A supply chain as a network of auctions. *Decision Support Systems and Electronic Commerce* 50, 176–190.

Nik-Khah, E. (2005). *Designs on the Mechanism: Economics and the FCC Auctions*. Ph. D. thesis, University of Notre Dame, Notre Dame, Indiana, USA.

Niu, J. (2011). *Automated Auction Mechanism Design with Competing Marketplaces*. Ph. D. thesis, The Graduate School and University Center, The City University of New York, New York, NY, USA.

Niu, J., K. Cai, S. Parsons, M. Fasli, and X. Yao (2012). A grey-box approach to automated mechanism design. *Electronic-Commerce Research and Applications* 11(1). (Forthcoming).

Niu, J., K. Cai, S. Parsons, E. Gerding, and P. McBurney (2008). Characterizing effective auction mechanisms: Insights from the 2007 TAC Mechanism Design Competition. In Padgham, Parkes, Müller, and Parsons (Eds.), *Proceedings of the Seventh International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Estoril, Portugal, pp. 1079–1086.

Niu, J., K. Cai, S. Parsons, P. McBurney, and E. Gerding (2010). What the 2007 TAC Market Design Game tells us about effective auction mechanisms. *Journal of Autonomous Agents and Multiagent Systems* 21(2), 172–203.

Niu, J., K. Cai, S. Parsons, and E. Sklar (2006). 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 (AAMAS 2006)*, Hakodate, Japan, pp. 1143–1150.

Niu, J., K. Cai, S. Parsons, and E. Sklar (2007). Some preliminary results on competition between markets for automated traders. In *Proceedings of AAAI-07 Workshop on Trading Agent Design and Analysis (TADA-07)*.

Niu, J., A. Mmoloke, P. McBurney, and S. Parsons (2009). CATP: a communication protocol for CAT games (version 2.0). Technical Report ULCS-09-012, Department of Computer Science, University of Liverpool, Liverpool, UK. Version 2.0.

Pardoe, D., P. Stone, M. Saar-Tsechansky, T. Keskin, and K. Tomak (2010). Adaptive auction mechanism design and the incorporation of prior knowledge. *INFORMS Journal on Computing* 22(3), 353–370.

Petric, A., V. Podobnik, A. Grguric, and M. Zemljic (2008). 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.

Phelps, S., P. McBurney, and S. Parsons (2010). Evolutionary mechanism design: a review. *Autonomous Agents and Multi-Agent Systems* 21(2), 237–264.

Robinson, E., P. McBurney, and X. Yao (2010). How specialised are specialists? Generalisation properties of entries from the 2008 and 2009 TAC Market Design competitions. In *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pp. 178–194. Springer.

Robinson, E. R. (2011). *Resource Allocation via Competing Marketplaces*. Ph. D. thesis, University of Birmingham, UK, Birmingham, UK.

Shi, B. (2011). *Analysis and Design of Competing Double Auction Marketplaces*. Ph. D. thesis, University of Southampton, UK, Southampton, UK.

Shi, B., E. Gerding, P. Vytelingum, and N. Jennings (2012). An equilibrium analysis of market selection strategies and fee strategies in competing double auction marketplaces. *Journal of Autonomous Agents and Multi-Agent Systems*. (Forthcoming).

Sohn, J., A. Kwasnica, and T. Mullen (2010). Human traders across multiple markets: Attracting intra-marginal traders under economic experiment. In *ACM EC 2010 Workshop on Trading Agent Design and Analysis*, Cambridge, MA, USA.

Sohn, J., S. Lee, and T. Mullen (2009a). Attracting intra-marginal traders across multiple markets. In *IJCAI-09 Workshop on Trading Agent Design and Analysis*, Pasadena, California.

Sohn, J., S. Lee, and T. Mullen (2009b). Impact of misalignment of trading agent strategy under a multiple market. In *Proceedings of the First Conference on Auctions, Market Mechanisms and Their Applications*, pp. 40–54. Springer.

Stavrogiannis, L. C. and P. A. Mitkas (2009). CAT 2008 post-tournament evaluation: The Mercator’s perspective. In *IJCAI-09 Workshop on Trading Agent Design and Analysis*, Pasadena, California.

Stavrogiannis, L. C. and P. A. Mitkas (2010). Mertacor 2009 demystified: Strategies and guidelines for market design in TAC. In *ACM EC 2010 Workshop on Trading Agent Design and Analysis*, Cambridge, MA, USA.

Sutton, R. S. and A. G. Barto (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT press.

Vytelingum, P., I. A. Vetsikas, B. Shi, and N. R. Jennings (2008). IAMwildCAT: The winning strategy for the TAC Market Design Competition. In *Proceedings of 18th European Conference on Artificial Intelligence*, Patras, Greece, pp. 428–434.

Wellman, M. P., A. Greenwald, and P. Stone (2007). *Autonomous Bidding Agents: Strategies and Lessons from the Trading Agent Competition*. Cambridge, MA, USA: MIT Press.

Wurman, P. R., W. E. Walsh, and M. P. Wellman (1998). Flexible double auctions for electronic commerce: Theory and implementation. *Decision Support Systems* 24(1), 17–27.