Trading Agents Competing: Performance, Progress, and Market Effectiveness

Michael P. Wellman, Shih-Fen Cheng, Daniel M. Reeves, and Kevin M. Lochner,
University of Michigan Artificial Intelligence Laboratory

Analyzing data from Trading Agent Competitions can quantify TAC travel market effectiveness in terms of allocative efficiency and identify opportunities for further gains from trade.

The annual Trading Agent Competition offers agent designers a forum for evaluating programmed trading techniques in a challenging market scenario. TAC aims to spur research by enabling researchers to compare techniques on a common problem and build on each other's ideas.[1] A fixed set of assumptions and environment settings facilitates communication of methods and results.

As a multiyear event, TAC lets researchers observe trading agents' progress over time, in effect accelerating the evolution of an adapted population of traders. Given all the participant effort invested, it is incumbent on us to learn as much from the experience as possible. After three years of TAC, we're ready to examine where we stand. To do this, we used data from actual TAC tournaments and some postcompetition experimentation. We based our analysis almost entirely on outcomes (profits and allocations), with very little direct accounting for specific agent techniques.

TAC 02 tournament results

Although we agree with those who caution against focusing excessively on ranked results in research competitions,[2] tournament results provide important information about agent quality. (See the related sidebar for a description of the TAC market game.) Agents presumably act to maximize expected score, so, all else being equal, an increased score reflects an improved agent. If several agents improve, however, this might not lead to higher total scores for those agents. Whereas certain improvements unambiguously increase total agent surplus (for example, fewer wasted flights and better entertainment allocation), others could reduce the value that agents retain (smarter agents could more effectively compete away the consumer surplus) or even the total system surplus (for example, deadweight loss due to strategic behavior).

Table 1 presents average scores for the 16 agents that played in the TAC 02 final and semifinal rounds. The fifth column shows final-round scores adjusted to correct for the fact that some agents get "easier" assignments than others, calculated according to the formula we developed for TAC 01.[3]

Tracking benchmark performance levels by keeping some agents constant lets us measure progress over time. For example, in the Conference on Automated Deduction's Automated Theorem Proving competition series,[4] a given year's best systems typically enter unchanged in the next year's event (along with improved versions, of course), which lets observers directly measure relative performance over time. In a game setting, where other agents comprise the environment, it isn't strictly fair to judge an agent with respect to a different field. Nevertheless, we can learn much from observing the implications of such transplants.

The 2001 tournament[3] included two calibrating agents in the seeding round: ATTac and dummy_buyer. ATTac-2000[5] was the highest-scoring agent from the TAC 00 finals. To account for the rule changes between TAC 00 and TAC 01, the ATTac team modified the agent with a one-line edit causing it to place all of its bids before the first hotel closure rather than during the game's last minute. We also included dummy_buyer, the agent the Michigan TAC team provided in 2001 to

1094-7167/03/$17.00 © 2003 IEEE
Published by the IEEE Computer Society

The Classic Game

The "classic" TAC game—distinguished from newer variants such as the supply chain game introduced for TAC 03[1]—presents a shopping task where traders assemble flights, hotels, and entertainment into trips for a set of eight probabilistically generated clients. The game describes clients by their preferred arrival and departure days (*pa* and *pd*), the premium (*hp*) they are willing to pay to stay at the "Towers" (T) hotel rather than "Shanties" (S), and their respective values (e_1, e_2, e_3) for three different entertainment events. The agents try to maximize the trips' value for their clients, net of their expenditures in the travel goods market. The agents exchange three categories of goods through three distinct market mechanisms.

Flights

A feasible trip includes air transportation both ways, comprising an inflight day i and outflight day j, where $1 \le i < j \le 5$. Flights in and out each day are sold independently, at prices determined by a stochastic process. Each flight initially costs $\sim U\,[250, 400]$ and follows a random walk thereafter with an increasingly upward bias.

Hotels

Feasible trips must also include a room in one of the two hotels for each night of the client's stay. Each hotel has 16 rooms available each night; these are sold through ascending 16th-price auctions. Agents submit bids for various quantities, specifying the price offered for each additional unit. When the auction closes, the game allocates units to the 16 highest offers, with all bidders paying the price of the lowest winning offer. Each minute, the hotel auctions issue quotes, indicating the 16th (*ask*) and 17th highest (*bid*) prices among the currently active unit offers.[2] Starting at minute four, the game randomly selects one hotel auction to close, with the others remaining active and open for bids.

Hotel bidders are also subject to a "beat the quote" rule,[3] requiring any new bid to offer to purchase at least one unit at a price of *ask* + 1, and at least as many units at *ask* + 1 as the agent was previously winning at *ask*.

Entertainment

Agents receive an initial random allocation of entertainment tickets (indexed by type and day), which they can allocate to their own clients or sell to other agents through continuous double auctions.[4] The entertainment auctions issue *bid* and *ask* quotes representing the highest outstanding buy and lowest sell offer and remain open for buying and selling throughout the 12-minute game duration. A client can sell tickets it doesn't own but must pay a penalty of 200 per ticket for any short sales not covered by the game's end.

The game defines a feasible client trip r by an inflight day in_r, outflight day out_r, hotel type (H_r, 1 if T and 0 if S), and entertainment types (E_r, a subset of {1, 2, 3}). This trip's value is given by

$$v(r) = 1{,}000 - 100\left(\left|pa - in_r\right| + \left|pd - out_r\right|\right) + hp \cdot H_r + \sum_{i \in E_r} e_i$$

At the end of a game instance, the TAC server calculates the optimal allocation of trips to clients for each agent, given final holdings of flights, hotels, and entertainment. The agent's game score is its total client trip utility minus net expenditures in the TAC auctions.

TAC has certainly succeeded in spurring research. Over a dozen publications reporting on the competitions, specific agents, techniques employed, and analyses have appeared to date in archival journals, refereed conferences, and magazines. (See http://auction2.eecs.umich.edu/researchreport.html for a comprehensive list.) The TAC literature thus represents an uncommonly rich corpus of documentation on trading strategy and behavior for a particular complex environment. Many accounts include specific analyses or experiments involving agents from multiple developers, or variants on a particular agent inspired by techniques reportedly employed by others.[5–7] Such efforts augment entrants' anecdotal reports that each successive year they incorporate the previously presented lessons and approaches in new and improved agent designs.

See www.sics.se/tac for a list of TAC participant affiliations, team leaders, and preliminary round results. Complete game logs are available for this and previous TAC events. Amy Greenwald has also collected brief agent descriptions.[8]

References

1. N. Sadeh et al., "TAC-03: A Supply-Chain Trading Competition," *AI Magazine*, vol. 24, no. 1, 2003, pp. 92–94.
2. P.R. Wurman, M.P. Wellman, and W.E. Walsh, "Flexible Double Auctions for Electronic Commerce: Theory and Implementation," *Decision Support Systems*, vol. 24, 1998, pp. 17–27.
3. P.R. Wurman, M.P. Wellman, and W.E. Walsh, "A Parametrization of the Auction Design Space," *Games and Economic Behavior*, vol. 35, 2001, pp. 304–338.
4. D. Friedman and J. Rust, eds., *The Double Auction Market*, Addison-Wesley, 1993.
5. A. Greenwald, "Bidding under Uncertainty in Simultaneous Auctions," *Proc. Int'l Joint Conf. Artificial Intelligence Workshop on Trading Agent Design and Analysis*, AAAI Press, 2003.
6. P. Stone et al., "ATTac-2001: A Learning, Autonomous Bidding Agent," *Proc. Agent-Mediated Electronic Commerce IV*, LNCS 2153, Springer-Verlag, 2002.
7. M.P. Wellman et al., "Price Prediction in a Trading Agent Competition," *J. Artificial Intelligence Research*, vol. 19, 2003.
8. A. Greenwald, "The 2002 Trading Agent Competition: An Overview of Agent Strategies," *AI Magazine*, vol. 24, no. 1, 2003, pp. 83–91.

play in test games without a full slate of agents. Whereas most other teams modified their agents' behaviors between (and during) the qualifying and seeding rounds, the dummy remained unchanged. Not surprisingly, we observed substantial deterioration in the dummy's standing as the preliminary rounds progressed.

The 2002 tournament didn't explicitly insert calibrating agents, but 12 agents returning from 2001 provided some natural calibration. In particular, TAC 01's top-scoring agents, livingagents[6] and ATTac-2001, participated essentially unchanged in TAC 02. Table 1 shows that livingagents did quite

Table 1. The Trading Agent Competition 2002 seeded agents' average semifinal and final round scores.

Agent	Affiliation	Semifinal score (14 games)	Final score (32 games)	Adjusted final score
ATTac	AT&T Research and colleagues	H1: 3,137	—	—
cuhk	Chinese Univ. of Hong Kong	H2: 3,266	3,069	3,045
kavayaH	Oracle India	H1: 3,200	3,099	3,039
livingagents*	Living Systems	H1: 3,310	3,181	3,161
PackaTAC	N. Carolina State Univ.	H2: 3,250	—	—
PainInNEC	NEC Research and colleagues	H1: 2,193	—	—
RoxyBot	Brown Univ.	H2: 3,160	—	—
sics	Swedish Inst. of Computer Science	H2: 3,146	—	—
SouthamptonTAC	Univ. of Southampton	H1: 3,397	3,385	3,337
Thalis	Univ. of Essex	H2: 3,199	3,246	3,210
tniTac	Poli Bucharest	H1: 3,108	—	—
TOMAhack	Univ. of Toronto	H2: 2,843	—	—
tvad	Technion	H1: 2,724	—	—
UMBCTAC	Univ. of Maryland, Baltimore County	H1: 3,208	3,236	3,291
Walverine	Univ. of Michigan	H2: 3,287	3,210	3,277
whitebear	Cornell Univ.	H2: 3,324	3,413	3,479

* Missing two games adversely affected livingagents' score. Discounting these would have led to an average score of 3,393.

well, assuming we ignore the bug that caused it to skip two games. ATTac, the top scorer in the TAC 02 seeding rounds, was eliminated in the semifinals. According to team leader Peter Stone, this was possibly because of technical difficulties resulting from a computational environment change between the seeding and semifinal rounds. Another more substantive possibility is that 2002 preliminary-round prices (which ATTac used as training data) didn't sufficiently reflect the final-round prices. We suspect that the decrease in relative performance also reflects a general increase in other agents' competence. Interestingly, because livingagents benefits by operating with effective and adaptive agents,[3] it performs robustly as others in the field improve.

TAC 02's two top-scoring agents, whitebear and SouthamptonTAC,[7] also contended in TAC 01. These agents reportedly evolved from their 2001 designs, adopting refined game environment classifications[8] and undergoing extensive experimentation and parameter tuning.[9] Of the eight other repeat entries,

- Three (Thalis, sics, UMBCTAC) represent complete reimplementations of their TAC 01 forms by different agent designers.
- Two (RoxyBot, cuhk) represent significant redesigns by the same designers.
- Three (PainInNEC, BigRed, harami) incorporate incremental or unknown changes by the same or related designers.

Market efficiency

We can gauge agents' effectiveness by how well they allocate travel goods, in the aggregate, through their market interactions. This is an indirect measure at best because each agent seeks to maximize its own surplus, not that of the overall system (comprising all agents plus the TAC seller). Nevertheless, such a social-welfare analysis provides a benchmark and sheds light on resource allocation through an economy of interacting software agents.

TAC 02

We measure aggregate effectiveness by comparing actual TAC market allocations with ideal global allocations, centrally calculated assuming knowledge of all client preference information. Consider the total group of 64 clients and the set of available resources: 16 hotel rooms and eight entertainment tickets of each type per day (see the sidebar for more details). The global optimizer calculates a resource allocation that maximizes total client utility, net of flight expenditures assuming availability at their initial prices. We take initial prices to be the flights' relevant inherent cost (exogenously determined, independent of TAC agent demand), treating the expected stochastic increase during the game as a cost-of-decision delay that the idealized optimizer would avoid. The global optimization completely neglects hotel and entertainment prices, as these are endogenous to the TAC market. Monetary transfers affect surplus distribution across TAC buyers and sellers, but not the total amount of surplus. We formulate the optimization problem as an integer linear program and solve it using CPLEX.

Table 2 shows the average idealized net utility, per client, in the various TAC rounds as determined by global optimization. "TAC market" reports average net utility achieved in the TAC games (also neglecting hotel and entertainment expenditures, but counting actual flight payments).

The TAC market achieved 89 percent of the optimal value, on average, in the final round's 32 games—a steady improvement from the qualifying (67 percent), seeding (76 percent), and semifinal rounds (88 percent). All differences are significant ($p < 0.01$) except the small increment from semifinals to finals. (Henceforth, all assertions of statistical significance refer to the 0.01 level, unless otherwise specified.)

We can't easily assess this effectiveness in absolute terms, so we provide two benchmarks for comparison:

- *Uniform hotel and entertainment*. We distributed the hotel rooms and entertainment evenly across the eight agents, then optimized each agent's allocation to clients. This approach yielded 95.2 percent of the globally optimal value on average. (Allocation values significantly exceeded the market in every round.)
- *Uniform hotel and endowed entertainment*. The relative average value dropped to 85.4 percent when we distributed only the hotels, leaving agents with their original entertainment endowment. (This value exceeded market results in qualifying and seeding rounds, with the market performing significantly better in the finals.)

Perhaps surprisingly, simply dividing the goods uniformly achieves a high fraction of the available surplus—better than the market if the distribution includes entertainment. One reason is that the agents are *ex ante* symmetric, with independent and identically distributed clients. Hotels thus provide little potential gain from trade. Second, a direct allocation avoids the obstacles agents face in pursuing their allotments individually through the market. The risks of price uncertainty, exposure due to complementarities, and unknown hotel-closing patterns necessarily entail some loss in expected allocation quality. For example, agents have sufficient hotel room availability to obtain trips for all clients (albeit shortened from desired lengths) and, given a definite allocation, can optimize for clients accordingly. With uncertainty, the agents can plan for longer trips than are jointly feasible and thus wind up wasting flights, hoarding hotel rooms (to hedge), or resorting to suboptimal fallback trip options. A preliminary analysis indicated that this uncertainty causes much misallocation in TAC play. In future work, we'll focus on developing a precise quantitative characterization of loss sources.

Table 2. TAC 02 market efficiency compared to the global optimum.

Round	Games	Global optimization: idealized net utility	TAC market's actual net utility	Market efficiency (% of optimal)
Qualifying	390	618	415	67.0
Seeding	1,045	618	470	75.7
Semifinal	28	608	534	87.7
Final	32	609	542	89.1

Comparing agent profiles

Given that programmers actively debug and develop their agents during preliminary rounds, we can assume that agent competence improves in succeeding tournament rounds. Selecting the best performers for the semifinals and finals naturally amplifies this effect. Thus, the progressive market efficiency improvement we see in Table 2 coincides with individual agent progress. We performed the same global optimization analysis for the TAC 01 finals (24 games) and found a market efficiency of 85.7 percent. Though better than the TAC 02 qualifying and seeding rounds, the TAC 01 finalists didn't allocate resources as well as TAC 02 finalists ($p = 0.024$) or even semifinalists ($p = 0.097$). This indirectly (to the extent that overall market efficiency aligns with individual agent success) confirms our conjecture that the 2002 agents were, overall, more competent than their predecessors.

Comparing market effectiveness across different agent configurations also yields interesting results. Our initial explorations employed versions of Michigan's TAC 02 entry, Walverine.[10] In 173 games with Walverine playing all eight agent slots, the market achieved 88.5 percent efficiency, a result statistically indistinguishable from that of the TAC 02 finalist pool. Of course, Walverine, like the others, promotes its own profit, not overall efficiency per se.

In future work, we intend to evaluate additional agent configurations, including further Walverine variants and other groups' agents. Our analysis requires only data describing client preferences and final allocations, and we welcome any game logs that other researchers might submit for our efficiency analysis, covering whatever profiles of their own agent (and variants) they've investigated. We also invite participation in further experiments involving mixtures of agents from multiple groups. One interesting question (at least as another benchmark) is what level of overall efficiency agents can attain when designed with this objective in mind. We should be able to achieve at least the level of our uniform hotel and entertainment benchmark (95.2 percent), given the simplicity of constructing bidding policies that uniformly distribute all goods.

Entertainment trading efficiency

Entertainment trading proved amenable to separate analysis, with interesting results. The game initially distributes entertainment goods as endowments to the agents, which then exchange among themselves through continuous double auctions (CDAs) to reach a final allocation. Although agents value entertainment based on their choice of trip dates, we can characterize with reasonable accuracy the gains from trade specifically attributable to the TAC market's entertainment component.

To measure entertainment trading efficiency, we simply compare the trip utility's aggregate "fun bonus" component in the globally optimal allocation with actual TAC market results. Table 3 shows efficiency percentages for the various game sets (and repeats Table 2's overall efficiency results for reference).

Interestingly, the entertainment perfor-

Table 3. TAC entertainment market efficiency compared to the global optimum.

Round	Games	Entertainment market efficiency (% of optimum)	TAC market efficiency (% of optimum)
2002 qualify	390	71.1	67.0
2002 seeding	1,045	79.0	75.7
2002 semifinal	28	83.0	87.7
2002 final	32	85.3	89.1
2001 final	24	85.5	85.7
All Walverine	173	85.4	88.5
Nonshading Walverine	55	85.1	89.4
Shading equilibrium	470	85.6	89.2

mance in the TAC 02 and TAC 01 finals virtually matched, despite the significant improvement in overall market performance. This suggests that strategic progress focused on hotel and flight strategies, which agrees with entrants' reports of their concentrations of effort.

The entertainment performance data also helps us calibrate estimates of potential gains from agent improvements. The two benchmarks based on uniform hotel distributions (one with no entertainment trading and one with uniform entertainment allocation) define a potential trading gain to uniformity of approximately 60.5 per client (484 per agent) given the fixed uniform hotel allocation. In training Walverine's entertainment strategy,[10] we observed an average difference of 478 between a policy of not trading entertainment (average fun bonus 1,019) and that of a representative hand-coded strategy (that of livingagents, average fun bonus 1,497). By contrast, the average fun bonus in the global optimal allocation runs around 1,677.

The closeness between the observed gain from trade in training and from uniform entertainment distribution in the uniform-hotel case suggests that entertainment trading's remaining benefit equals the difference between uniform and optimal. (The TAC 02 finals fun bonus turned out lower than expected based on training observations for both Walverine and livingagents, and apparently for the rest of the field except for whitebear.[10]) We evaluated this by calculating a third benchmark based on global optimization of entertainment subject to a uniform allocation of hotels to agents. The result is 92.7 per agent greater than the value with uniform entertainment allocation. Although these measurements depend on the uniform hotel allocations, we believe the relative entertainment values will prove robust to any reasonable fixed hotel allocation.

Individual versus social welfare

As emphasized above, measuring market efficiency differs from measuring individual agents' effectiveness. Some agent competencies, such as choosing optimal trips and avoiding wasted flights and hotels, correspond directly to overall efficiency improvements. Others, such as reducing hotel demand to capture more seller surplus, tend to reward individual agents at the expense of social welfare. Understanding how this tradeoff operates in the TAC game sheds light on the relationship between market efficiency measures and agent performance assessment.

We can study this by identifying particular strategy components that we'd expect to detract from overall welfare and measuring the welfare loss between the socially and individually optimal policies. We consider behavior individually optimal if it is part of a Bayes-Nash equilibrium for the game. Although deriving BNE for TAC isn't remotely tractable, we've found it feasible to characterize restricted TAC BNE with respect to highly constrained subsets of allowable strategies.

Specifically, we investigated the strategic behavior of hotel bid *shading*, where the agent offers to buy rooms at prices lower than its marginal unit values. Walverine determines bid prices as part of a decision-theoretic optimization of expected surplus, which accounts for the probability of not obtaining a good even though its price is below the agent's valuation.[10] (We refer to this as *optimal bidding*, although the optimization embodies several simplifying assumptions. We suspect there's much room for improvement and are redesigning this component for future Walverine versions.) Although it benefits the individual agent by design, such shading detracts from market efficiency because the market generally doesn't receive faithful signals of goods' relative value to the various agents. (If all agents shaded proportionally, the relative offer prices would still provide relevant information to the market. In general, however, price reductions don't cancel out as such because the bidder's optimization includes many agent-specific contextual factors.)

Specifically, we investigated the strategic behavior of hotel bid shading, where the agent offers to buy rooms at prices lower than its marginal unit values.

To evaluate this effect, we defined a variant *nonshading* strategy that Walverine implemented with its optimal bidding procedure (shading) turned off. Agents in this version bid their true marginal values for every hotel room. We hypothesized that this would improve social welfare but sacrifice individual profits. Indeed, we found in a 55-game trial with all nonshading Walverines that the market achieved 89.4 percent efficiency—better than all Walverines ($p = 0.06$). The actual effect of varying strategy, however, generally depends on other agents' strategies. We therefore employed an evolutionary search approach to find a restricted BNE for this game.[11]

We began by running a series of TAC games, distributed over the nine possible profiles of shading and nonshading agents. Averaging the scores for each profile (adjusted for client preference favorability) yielded an expected payoff for each strategy in each profile. We then calculated a BNE for the restricted game using replicator dynamics[12] (Gambit wasn't able to find a symmetric BNE after some hours of CPU time, finding only the asymmetric equilibrium in which five agents shade and three do not). In particular, we identified a symmetric mixed-strategy BNE, where agents shade with probability 0.11 and refrain from shading with probability 0.89. This result applies specifically to Walverine's shading implementation; alternate shading policies incorporated in other agent strategies could produce varying outcomes.

The predominance of truthful bidding in equilibrium demonstrates this policy's advantages in a population with substantial shaders. It isn't dominant, however, so as the population approaches all nonshaders, shading offers benefits. In equilibrium, shading and nonshading agents have equal payoffs, each a best response to the given mixture. The average client-adjusted payoffs for all shading, all nonshading, and the BNE mixed strategy are 3,339, 3,155, and 3,209, with corresponding market efficiencies of 88.5 percent, 89.4 percent, and 89.2 percent. Playing the equilibrium strategy results in an average payoff gain of 53 per agent per game but a loss of 46 in social welfare compared to all nonshaders.

This analysis provides evidence for TAC traders' competence and progress within and between TAC tournaments, although because we use indirect measures (for example, measuring market efficiency rather than absolute agent performance), we can't draw definitive conclusions. We could base more compelling demonstrations of progress and competence on further calibration studies, systematic searches in strategy spaces, and attribution of allocation suboptimality among its many possible causes (for example, agent suboptimality and inherent risk—including cost of its rational management). Further benchmarks capturing less ideal conditions might prove useful in this regard.

Another natural question we haven't addressed is how well TAC agents fare compared to what human traders could do. We've found no evidence that humans would be particularly adept at a TAC-like trading task. One of the few studies comparing human and computer traders (in an abstract CDA scenario) didn't reflect very favorably on the humans.[13]

Acknowledgments

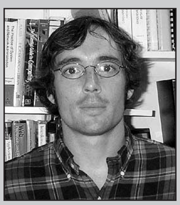
This is a revised and extended version of a paper appearing in the International Joint Conference on Artificial Intelligence 2003 Workshop on Trading Agent Design and Analysis. We benefited much from discussions with the rest of the Michigan TAC Team and from comments by the anonymous reviewers. National Science Foundation grants IIS-9988715 and IIS-0205435 supported this research in part.

The Authors

Michael P. Wellman is a professor of computer science and engineering and director of the Artificial Intelligence Laboratory at the University of Michigan. His research lies at the intersection of artificial intelligence and electronic commerce, with special emphasis on market games. He received his PhD in artificial intelligence from the Massachusetts Institute of Technology. He currently chairs the ACM Special Interest Group on E-Commerce (SIGecom). Contact him at the Univ. of Michigan Artificial Intelligence Lab., 1101 Beal Ave., Ann Arbor, MI 48109-2110; wellman@umich.edu.

Shih-Fen Cheng is a PhD candidate in industrial and operations engineering at the University of Michigan. His research interests include artificial intelligence and applying game-theoretic approaches to the modeling and solving of large-scale optimization problems. He received his BSE in mechanical engineering at the National Taiwan University. He is a member of the American Association for Artificial Intelligence. Contact him at the Univ. of Michigan Artificial Intelligence Lab., 1101 Beal Ave., Ann Arbor, MI 48109-2110; chengsf@umich.edu.

Daniel M. Reeves is a PhD candidate in computer science and engineering at the University of Michigan. His research interests involve AI in electronic commerce, in particular, the automatic generation of strategies for trading agents. He received his MSE in computer science and engineering from the University of Michigan. Contact him at the Univ. of Michigan Artificial Intelligence Lab., 1101 Beal Ave., Ann Arbor, MI 48109-2110; dreeves@umich.edu.



Kevin M. Lochner is a PhD candidate in computer science and engineering at the University of Michigan. His research interests include trading agents, auctions, game theory, and mechanism design. He received his MSE in computer science and engineering from the University of Michigan and his BSE in electrical engineering and economics from Duke University. Contact him at the Univ. of Michigan Artificial Intelligence Lab., 1101 Beal Ave., Ann Arbor, MI 48109-2110; klochner@umich.edu.

References

1. M.P. Wellman and P.R. Wurman, "A Trading Agent Competition for the Research Community," *Proc. Int'l Joint Conf. Artificial Intelligence Workshop on Agent-Mediated Electronic Trading*, AAAI Press, 1999.
2. P. Stone, "Multiagent Competitions and Research: Lessons from RoboCup and TAC, *RoboCup 2002: Robot Soccer World Cup VI*, LNAI 2752, Springer-Verlag, 2002.
3. M.P. Wellman et al., "The 2001 Trading Agent Competition," *Electronic Markets*, vol. 13, no. 1, 2003, pp. 4–12.
4. G. Sutcliffe, "The CADE-17 ATP System Competition," *J. Automated Reasoning*, vol. 27, 2001, pp. 227–250.
5. P. Stone et al., "ATTac-2000: An Adaptive Autonomous Bidding Agent, *J. Artificial Intelligence Research*, vol. 15, 2001, pp. 189–206.
6. C. Fritschi and K. Dorer, "Agent-Oriented Software Engineering for Successful TAC Participation," *Proc. 1st Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, ACM Press, 2002.
7. M. He and N.R. Jennings, "SouthamptonTAC: Designing a Successful Trading Agent," *Proc. 15th European Conf. Artificial Intelligence*, IOS Press, 2002, pp. 8–12.
8. M. He and N.R. Jennings, "SouthamptonTAC: An Adaptive Autonomous Trading Agent," *ACM Trans. Internet Technology*, vol. 3, 2003, pp. 218–235.
9. I.A. Vetsikas and B. Selman, "A Principled Study of the Design Tradeoffs for Autonomous Trading Agents," *Proc. 2nd Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, ACM Press, 2003, pp. 473–480.
10. S.-F. Cheng et al., "Walverine: A Walrasian Trading Agent," *Proc. 2nd Int'l Joint Conf. Autonomous Agents and Multiagent Systems*, ACM Press, 2003, pp. 465–472.
11. M.P. Wellman et al., "Exploring Bidding Strategies for Market-Based Scheduling," *Proc. 4th ACM Conf. Electronic Commerce*, ACM Press, 2003, pp. 115–124.
12. P. Taylor and L. Jonker, "Evolutionary Stable Strategies and Game Dynamics," *Mathematical Biosciences*, vol. 40, 1978, pp. 145–156.
13. R. Das et al., "Agent-Human Interactions in the Continuous Double Auction," *Proc. 17th Int'l Joint Conf. Artificial Intelligence*, AAAI Press, 2001, pp. 1169–1176.