Multiagent Competitions and Research: Lessons from RoboCup and TAC

Peter Stone

Department of Computer Sciences
The University of Texas at Austin
Austin, Texas 78712-1188
pstone@cs.utexas.edu
http://www.cs.utexas.edu/~pstone

Abstract. This paper compares and contrasts two recent series of competitions in which multiple agents compete directly against one another: the robot soccer world cup (RoboCup) and the trading agent competition (TAC). Both of these competitions have attracted large numbers of competitors and have motivated important research results in artificial intelligence. Based on extensive personal experiences, both as a participant and as an organizer, this paper reflects upon and characterizes both the benefits and the hazards of competitions with respect to academic research.

1 Introduction

Competitions are becoming increasingly prevalent in the research world. For one example, the annual Loebner competition [13] challenges entrants to create a computer program whose verbal responses in conversation are judged to be most "human." For another example, the recent planning systems competition [15] compared the performance of AI planning systems in a variety of planning domains.

In both the Loebner and planning competitions, the judging and/or scoring is done independently for each program: an entrant's score does not depend on the behavior of the other entrants. However, there have also been several competitions in which the agents *do* interact directly. Examples include Axelrod's iterated prisoner's dilemma (IPD) tournament from the late 1970's [3]; the Santa Fe double auction tournament in the late 1980's [16]; and the recent RoShamBo (rock-paper-scissors) programming competition [4]. All three of these competitions led to interesting results despite the fact that entered programs faced very limited sensation and action spaces in domains that have been well-studied and understood in isolation (non-iterated).

This paper compares and contrasts two recent series of competitions in which—like IPD, double auctions, and RoShamBo—agents must interact directly with one another, and in which the substrate domains are much more complicated and difficult to analyze in isolation: the robot soccer world cup (RoboCup) and the trading agent competition (TAC). In both cases, the agents face complex sensations with a good deal of hidden state and, for all intents and purposes, continuous action spaces. The success of agent strategies depends a great deal on the strategies of the other competitors.

Both of these competitions have attracted large numbers of competitors and motivated important research results in artificial intelligence and robotics. However, along with the many benefits, competitions also bring with them some potential hazards. Based on extensive personal experiences within RoboCup and TAC, both as a successful participant and as an organizer, this paper reflects upon and characterizes the benefits and hazards of competitions with respect to the research world.

The remainder of the paper is organized as follows. Section 2 gives a brief overview of the RoboCup domain. Section 3 gives a more detailed introduction to TAC. Section 4 compares the relevant features of these two domains and Section 5 lays out the potential benefits and hazards of these and other competitions. Section 6 concludes.

2 RoboCup

RoboCup is an international initiative designed to encourage research in the fields of robotics and artificial intelligence, with a particular focus on developing cooperation between autonomous agents in dynamic multiagent environments. It uses the game of soccer as the main underlying testbed. A long-term grand challenge posed by RoboCup is the creation of a team of humanoid robots that can beat the best human soccer team by the year 2050.

The first international RoboCup competition was held from August 23–29, 1997 in Nagoya, Japan. It involved 35 teams from 12 countries. By the 3rd RoboCup in 1999, there were 90 teams from 30 countries and RoboCup continues to grow in popularity, now including many regional events as well as the annual international championships.

Currently, RoboCup includes four robotic soccer competitions (simulation, small-size, mid-size, and legged), and two disaster rescue competitions (simulation and real-robot). A humanoid robotic soccer competition is also planned for the near future. Each of these competitions has its own format, but generally they are run over the course of about a week, beginning with round-robins and culminating with elimination tournaments. Competitors are all responsible for bringing their own robots and/or software to the competition site.

Some commonalities across the different soccer leagues are that they are run in *dynamic, real-time, distributed*, *multiagent* environments with both *teammates* and *adversaries*. In general, there is *hidden state*, meaning that each agent has only a partial world view at any given moment. The agents also have *noisy sensors and actuators*, meaning that they do not perceive the world exactly as it is, nor can they affect the world exactly as intended. In addition, the perception and action cycles are *asynchronous*, prohibiting the traditional AI paradigm of using perceptual input to trigger actions. *Communication* opportunities are limited; and the agents must make their decisions in *real-time*. These italicized domain characteristics combine to make robotic soccer a realistic and challenging domain.

For the purposes of this forum, I assume a general familiarity with RoboCup. For complete details on the competitions in the different leagues, see the continuing series of RoboCup books [11, 2, 23, 19].

3 Trading Agent Competition

The first Trading Agent Competition (TAC) was held from June 22nd to July 8th, 2000, organized by a group of researchers and developers from the University of Michigan and North Carolina State University [26]. Their goals included providing a benchmark problem in the complex and rapidly advancing domain of e-marketplaces [7] and motivating researchers to apply unique approaches to a common task. TAC-00 included 16 agents from 6 countries. Building on the success of TAC-00, TAC-01 included 19 agents from 9 countries. The precise rules and tournament structures changed slightly from TAC-00 to TAC-01. The details in this section reflect the state of affairs as of TAC-01.

TAC has been motivated in some ways by RoboCup. Although it is based on a fundamentally different problem—auction bidding as opposed to playing soccer, it is also a multiagent competition. The organizers explicitly took some cues from RoboCup in planning out the competition [25]. This section gives a detailed overview of the TAC domain.

One key feature of TAC is that it requires *autonomous bidding agents* to buy and sell *multiple interacting goods* in auctions of different types. Another key feature is that participating agents compete against each other in preliminary rounds consisting of many games leading up to the finals. Thus, developers change strategies in response to each others' agents in a sort of escalating arms race. Leading into the day of the finals, a wide variety of scenarios are generally possible. A successful agent needs to be able to perform well in any of these possible circumstances.

A TAC game instance pits 8 autonomous bidding agents against one another. Each TAC agent is a simulated travel agent with 8 clients, each of whom would like to travel from TACtown to Tampa and back again during a 5-day period. Each client is characterized by a random set of preferences for the possible arrival and departure dates; hotel rooms (Tampa Towers and Shoreline Shanties); and entertainment tickets (alligator wrestling, amusement park, and museum). In order to obtain utility for a client, an agent must construct a travel package for that client by purchasing airline tickets to and from TACtown and securing hotel reservations; it is possible to obtain additional utility by providing entertainment tickets as well. A TAC agent's score in a game instance is the difference between the sum of its clients' utilities for the packages they receive and the agent's total expenditure.

TAC agents buy flights, hotel rooms and entertainment tickets in different types of *auctions*. The TAC server, running at the University of Michigan, maintains the markets and sends price *quotes* to the agents. The agents connect over the Internet and send bids to the server that update the markets accordingly and execute transactions.

Each game instance lasts 12 minutes and includes a total of 28 auctions of 3 different types.

Flights (8 auctions): There is a separate auction for each type of airline ticket: flights to Tampa (*inflights*) on days 1–4 and flights from Tampa (*outflights*) on days 2–5. There is an *unlimited* supply of airline tickets, and their *ask price* changes every 24–32 seconds by from $\$-10$ to $\$x$. x increases linearly over the course of a game from 10–y, where $y \in [10, 90]$. y is independent for each auction, and is unknown to the bidders. In all cases, tickets are priced between \$150 and \$800. When the server receives a bid at or above the ask price, the transaction is *cleared immediately* at the ask price. *No resale* of airline tickets is allowed.

Hotel Rooms (8): There are two different types of hotel rooms—the Tampa Towers (TT) and the Shoreline Shanties (SS)—each of which has 16 rooms available on days 1–4. The rooms are sold in a 16th-price *ascending* (English) auction, meaning that for each of the 8 types of hotel rooms, the 16 highest bidders get the rooms at the 16th highest price. For example, if there are 15 bids for TT on day 2 at \$300, 2 bids at \$150, and any number of lower bids, the rooms are sold for \$150 to the 15 high bidders plus one of the \$150 bidders (earliest received bid). The ask price is the current 16th-highest bid. Thus, agents have no knowledge of, for example, the current highest bid. New bids must be higher than the current ask price. *No bid withdrawal* and *no resale* is allowed, though the price of bids may be lowered provided the agent does not reduce the number of rooms it would win were the auction to close. Transactions only *clear when the auction closes*. One randomly-chosen hotel auction closes at minutes 4–11 of the 12-minute game. Quotes are only changed on the minute.

Entertainment Tickets (12): Alligator wrestling, amusement park, and museum tickets are each sold for days 1–4 in *continuous double auctions*. Here, agents can *buy and sell* tickets, with transactions *clearing immediately* when one agent places a buy bid at a price at least as high as another agent's sell price. Unlike the other auction types in which the goods are sold from a centralized stock, each agent starts with a (skewed) random endowment of entertainment tickets. The prices sent to agents are the *bid-ask spreads*, i.e., the highest current bid price and the lowest current ask price (due to immediate clears ask price is always greater than bid price). In this case, *bid withdrawal* and *ticket resale* are both permitted. Each agent gets blocks of 4 tickets of 2 types, 2 tickets of another 2 types, and no tickets of the other 8 types.

In addition to unpredictable market prices, other sources of variability from game instance to game instance are the client profiles assigned to the agents and the random initial allotment of entertainment tickets. Each TAC agent has 8 clients with randomly assigned travel preferences. Clients have parameters for ideal arrival day, *IAD* (1–4); ideal departure day, *IDD* (2–5); hotel premium, *HP*

(\$50–\$150); and entertainment values, EV (\$0–\$200) for each type of entertainment ticket.

The utility obtained by a client is determined by the travel package that it is given in combination with its preferences. To obtain a non-zero utility, the client must be assigned a *feasible* travel package consisting of an arrival day AD with the corresponding inflight, departure day DD with the corresponding outflight, and hotel rooms of the *same type* (TT or SS) for each day d such that $AD \leq d < DD$. At most one entertainment ticket can be assigned for each day $AD \leq d < DD$, and no client can be given more than one of the same entertainment ticket type. Given a feasible package, the client's utility is defined as

$$1000 - \mathit{travelPenalty} + \mathit{hotelBonus} + \mathit{funBonus}$$

where

- $\mathit{travelPenalty} = 100(|AD - IAD| + |DD - IDD|)$
- $\mathit{hotelBonus} = HP$ if the client is in the TT, 0 otherwise.
- $\mathit{funBonus}$ = sum of relevant EV's for each entertainment ticket type assigned to the client.

A TAC agent's final score is simply the sum of its clients' utilities minus the agent's expenditures. Throughout the game instance, it must decide what bids to place in each of the 28 auctions. At the end of the game, it must submit a final allocation of purchased goods to its clients.

The client preferences, allocations, and resulting utilities from an example game are shown in Tables 1 and 2.

Client	IAD	IDD	HP	AW	AP	MU
1	Day 2	Day 5	73	175	34	24
2	Day 1	Day 3	125	113	124	57
3	Day 4	Day 5	73	157	12	177
4	Day 1	Day 2	102	50	67	49
5	Day 1	Day 3	75	12	135	110
6	Day 2	Day 4	86	197	8	59
7	Day 1	Day 5	90	56	197	162
8	Day 1	Day 3	50	79	92	136

Table 1. One agent's preferences from a sample game. AW, AP, and MU are EVs for alligator wrestling, amusement park, and museum respectively.

For full details on the design and mechanisms of the TAC server and TAC game, see `http://tac.eecs.umich.edu`.

TAC-01 was organized as a series of four competition phases, starting on September 10, 2001 and culminating with the semifinals and finals on October 14, 2001 at the EC-01 conference in Tampa, Florida. First, the qualifying round, consisting of about 270 games per agent, served to select the 16 agents that would

Client	*AD*	*DD*	Hotel	Entertainment	Utility
1	Day 2	Day 5	SS	MU4	1175
2	Day 1	Day 2	TT	AW1	1138
3	Day 3	Day 5	SS	MU3, AW4	1234
4	Day 1	Day 2	TT	None	1102
5	Day 1	Day 2	TT	AP1	1110
6	Day 2	Day 3	TT	AW2	1183
7	Day 1	Day 5	SS	AP2, AW3, MU4	1415
8	Day 1	Day 2	TT	MU1	1086

Table 2. The client allocations and utilities from the agent and game shown in Table 1. Client 1's "MU4" under "Entertainment" indicates museum on day 4.

participate in the semifinals. Second, the seeding round, consisting of about 315 games per agent, was used to divide these agents into two groups of eight. After the semifinals, consisting of 11 games per agent, on the morning of October 14th, four teams from each group were selected to compete in the 24-game finals during that same afternoon.

4 Comparisons

The RoboCup and TAC domains have many similarities when considered from a research perspective. For example, agents in both domains must deal with a dynamic, real-time environment with hidden state and unpredictable adversaries. In most RoboCup leagues, the team must be distributed among several agents[1], and in TAC, some of the agents have used a distributed approach (for example using one agent to bid on hotels and another to bid on entertainment tickets).

Another key similarity between the two domains is that they are complex enough to prohibit any realistic attempt to solve them from a game theoretic perspective, and there is very little chance that strategies will emerge that can be described as in "equilibrium" in any sense. Rather, to be successful in TAC or RoboCup, an agent must be robust to a wide variety of opponents or economies. Ideally, an agent should be able to adapt on-line as opponents and environments change.

There are also many differences between the two domains. In many ways, RoboCup introduces additional research challenges that are not present in TAC, such as sensor and actuator noise, communication among teammates, and asynchronous sensing and acting: TAC agents can operate by choosing a set of actions (bids) every time they receive a new set of sensations (prices).

On the other hand, One feature of TAC not present in RoboCup is that the agents play against each other over an extended period leading up to the finals. In both TAC-00 and TAC-01, the competitors learned about each others' strategies in the weeks leading up to the finals, and made many adjustments as a result. For example, in TAC-00, only 14% of the agents were using a particularly

[1] The small-size robot league allows for centralized control of all 5 robots.

effective (in isolation) high-bidding strategy during the qualifying round; by the finals 58% of the agents were using this strategy [21]. In RoboCup there is a good deal of technology exchange from year to year. But within a single year, teams tend to meet each other for the first time during the actual competition. Table 4 summarizes some of the characteristics of TAC and RoboCup.

	TAC	RoboCup
Dynamic	+	+
Real-time	+	+
Multiagent	+	+
Hidden state	+	+
Adversaries	+	+
Teammates	—	+
Noisy sensors	—	+
Noisy actuators	—	+
Noisy actuators	—	+
Asynchronous	—	+
Communication	—	+
Distributed	(+)	+
Repeated play	+	—

Table 3. Some comparisons between the TAC and RoboCup domains.

Another interesting relationship between TAC and RoboCup is that they have proven to be attractive as research domains to many of the same people. In TAC-01, at least 6 of the 19 entries either involved the same institution or some of the same people as previous RoboCup entries[2]. In fact, the top two teams from the RoboCup-99 simulator competition and from TAC-01 involved two of the same people on opposing teams[3].

5 Lessons Learned from Competitions

Competitions have the potential to accelerate scientific progress within specific domains. However, there are also many potential hazards that can render them detrimental to progress.

Both RoboCup and TAC are primarily research initiatives. As such, their primary goals are to help advance the state of the art. They have certainly done so by providing new and challenging domains for studying issues within robotics and AI, such as “design principles of autonomous agents, multiagent

[2] The teams from AT&T Labs - Research, Carnegie Mellon University, Cornell University, Swedish Institute of Computer Science, and the University of Essex.

[3] Klaus Dorer from magmaFreiburg (RoboCup-99) [6] and livingagents (TAC-01) [9]; and Peter Stone from CMUnited-99 (RoboCup-99) [22] and ATTac (TAC-01) [17].

collaboration, strategy acquisition, real-time reasoning and planning, intelligent robotics, sensor-fusion, and so forth" [12].

However, the domains exist without the competitions. In this section, I examine the potential hazards and potential benefits of having periodic large-scale competitions, drawing on my experiences as a participant and organizer. I operate under the premise that scientific progress (as opposed to, for example, entertainment) is the primary goal.

I start by examining the potential hazards of competitions; then I point out the potential benefits. As many potential hazards and benefits are quite similar, it is up to the participants and organizers to sway the balance towards the benefits.

5.1 Hazards

There are many potential hazards to scientific progress involved in holding organized competitions. However, many can be avoided through careful organization of the competitions along with an engineered social climate within the community. Here, I list the possible hazards while, where possible, indicating how RoboCup and TAC have tried to avoid them.

Obsession with winning. One of the most obvious potential hazards of competitions is that people try to win them at the expense of all else, including science. Especially if there are monetary prizes involved, many people will focus only on winning and there is a potential incentive to keep successful techniques secret from year to year. RoboCup and TAC both do their best to avoid this hazard by not awarding any monetary prizes. Instead, the winners are rewarded with opportunities to disseminate their research via invited publications. In addition, in RoboCup, "scientific challenge" awards are given to teams who, in the opinions of the organizers, have demonstrated the best scientific contributions in their teams. In comparison with the competition winners, winners of these scientific challenge awards are given equal, if not greater, status at the awards ceremonies and within the community. Thus, there is explicit incentive given to deemphasize winning in favor of focusing on scientific contributions. Nonetheless, competitive spirit can easily take over.

Domain-dependent solutions. Another potential hazard of competitions, particularly within complex domains, is that it can be difficult to avoid getting bogged down in the low-level details of the domain. If the competition is to serve scientific interests, the winning solutions should be ones that are generally applicable beyond the particular domain in question. Of course, it is impossible to avoid requiring some domain-dependent solutions. However, while necessary, they should not be sufficient to produce a winning team. One way to encourage an emphasis on high-level, generalizable solutions is to repeat the same competition several times. While the first iteration is likely to be won by the best domain-dependent solution, subsequent events are more likely to find several teams using the same low-level approach that has already been proven effective. Then the difference among the teams will be more at the general levels. For example, at RoboCup-97, the winning teams in both the simulator and small-robot competitions were the ones that had the best low-level sensing and acting capabilities. However at RoboCup-98, there were several teams with similar low-level capabilities. Similarly, the biggest differences among agents in TAC-00 were their approaches to the TAC-specific allocation sub-problem [20], while in TAC-01 many of the agents were solving it optimally by building upon the previous year's published approaches. Instead, the crucial differences were at the level of strategic reasoning using techniques that are not limited to the specific domains.

Cost escalation. Especially in the robot competitions, there is the potential to have increasingly expensive solutions. If an expensive technology provides a significant advantage at one competition, then it might become a prerequisite for success in future years. If the expense is prohibitive to academic researchers, then the competition could die out. This issue has not yet been addressed in RoboCup. One possible solution would be to require that all teams use a common hardware platform, restricting the differences to the software. In fact, the RoboCup legged robot competition uses this approach as the only robots meeting the competition specifications were the Sony legged robots [24]. However, in general, this is not a satisfactory approach for RoboCup given that some of the interesting research issues are in the creation of the hardware itself. Another possible solution would be to enforce cost limits on entrants. However, such a restriction would be very difficult to define and enforce adequately. Cost escalation may become a serious issue for RoboCup in the near future.

Barrier to entry. As a competition repeats from year to year, it is natural that the people who have been involved in the past have an advantage over new-comers. As time goes on, this effect can magnify to the point that new-comers can never hope to compete meaningfully: the barrier to entry becomes too high. For example, in the world of computer chess, the leaders in the field invested large amounts of time and money building specialized hardware expressly for the purpose. It became virtually impossible for a new-comer to get up to speed in a reasonable amount of time. One reason for this effect was that the rules of chess are well-defined and unchanging: a successful approach in one competition is likely to remain successful even if left unchanged. One way around this effect is to gradually change the rules from year to year in order to make them slightly more challenging. For example, from the first year to the second year, the TAC competition changed from having all of the hotel auctions close at the end of the game to having them close randomly over the course of the game. Thus the previous year's competitors had to address an important new challenge at the same time as the new-comers. The barrier to entry can also be lowered considerably if competitors make portions of their code available as has happened consistently in the RoboCup simulation league.

Restrictive rules. While it is important to have well-defined rules for competitions, there is a potential to discourage research innovations via these rules. Especially for competitions involving robots, it is difficult to create rules that have no loopholes but are not overly restrictive. Over the years, RoboCup competitions have been run with varying degrees of specificity in the rules. Since the first year, the simulator league has always included a general "unsportsmanlike" clause, generally prohibiting anything that is not "in the spirit of the competition." Similarly, the TAC organizers reserved "the right to disqualify agents violating the spirit of fair play." The RoboCup small-robot league, on the other hand, has tended towards precise, completely-specified rules. While the former approach has the potential to lead to some heated arguments, my experience is that it is the best from a research perspective since it discourages participants from focusing on exploiting minute aspects of the rules.

Invalid evaluation conclusions. There is the potential at competitions to conclude that if agent (or team) A beats team B, then all of the techniques used by team A are more successful than those used by team B. However, this conclusion is invalid. Unless the agents are identical except in one respect, no individual aspect of either agent can conclusively be credited with or blamed for the result. Indeed, I have been involved on teams that won several of the competitions described above, but we do not present the results of these competitions as evaluations of any of the contributions of our research other than the agents as a whole. Instead, we conduct extensive controlled experiments to validate our research contributions [18, 21, 17].

5.2 Benefits

While there are many potential hazards to holding competitions, there are also many potential benefits. Here I list the possible benefits, again illustrating them with specific examples from RoboCup and TAC whenever possible.

Research Inspiration. While one potential hazard of competitions stemming from peoples' competitive spirit is an obsession with winning, a related benefit is that competitions are a great source of research inspiration. Several research innovations have been the direct result of preparations for one of the above competitions. While they started as innovative solutions to challenging specific problems, participants were then able to abstract their contributions into general frameworks. The natural desire to win is a strong motivation to create a good team by solving the challenging aspects of the domain.

Deadlines for creating complete agents. Competitions create hard deadlines for the creation of *complete* working systems. In order to compete, it is not sufficient for any one component of the system to be operational. Therefore, entrants must confront the challenging issues of "closing the loop," i.e. getting all components working from sensing, to acting, to strategic reasoning. They must create complete agents. No matter how sophisticated a team's high-level strategic reasoning, if it does not solve the low-level issues, some other team will easily win. Our experience has been that these deadlines have forced us to solve difficult holistic problems that we might have otherwise overlooked: these problems have been a source of research inspiration for us.

Common platform for exchanging ideas. Competitions can bring together a group of people who have all tried to solve the same problems in the same domain. Unlike in many research communities, there is a common substrate system and a common language among participants. For example, in the planning community, researchers use a wide variety of planning systems, each with its own properties and idiosyncrasies, sometimes making it difficult to directly compare approaches and technique. Indeed, in a recent planning competition one main challenge was finding the commonalities and compatibilities among different planning representations [15]. In RoboCup and TAC, on the other hand, everyone implements their ideas in the same underlying architecture. Consequently, it is relatively easy to compare the various systems.

Continually improving solutions. When holding repeated competitions with the same platform, there is likely to be a continual improvement in solutions from event to event. All entrants know that in order to have a chance of winning a competition, they must be able to outperform the previous champion. Therefore, they are motivated to find some method of improving over the previous solutions. Of course, this benefit only applies if the same, or similar, rules are used as the basis for competition year after year. For example, in the AAAI robot competitions [1], there are some new tasks to be solved every year (in recent years there have also been some repeated tasks). While the new tasks encourage new entrants, there is no basis for directly measuring improvement from year to year.

Excitement for students at all levels. The inherent excitement of the RoboCup and TAC competitions encourages students at all levels to become involved in serious research. Competition entries often come from large teams of professors, graduate students, and undergraduates working together. By encouraging more people to become involved in research, the competitions can speed up progress. In addition, the competitions are ideal for undergraduate and graduate classes. There have been several courses around the world that have culminated in either simulated robotic soccer or trading agent competitions[4]. From all accounts, students in these classes have genuinely *enjoyed* putting in a good deal of time and effort to create their agents, and have learned a lot in the process.

Wide pool of teams created. After each competition, all of the entrants have created agents capable of performing in the given domain. If these agents are made available in some way, they can subsequently be used for controlled testing of research contributions. For example, in order to test technique x that is a single aspect of one's agent (or team), one could play the agent

[4] For instance, I recently taught a course using the RoboCup simulator (`http://cs.nyu.edu/courses/fall01/G22.3033-012/index.htm`). For a recent class that used TAC, see `http://ecommerce.ncsu.edu/csc513/`.

against another agent first with technique x active, and then without, thus establishing the effects of technique x. While such testing could be done against any agent, it is often up to the researchers themselves to create the team against which to test. As a result the comparison is often done against a trivial or simple team: a "straw-man." The competition can provide several teams against which to test, each of which is the result of serious effort by an independent group of researchers.

Generate realistic economies. Another related benefit of competitions is that realistic pools of teams, or economies in the case of TAC, can be studied. Presumably there are many groups in the financial industry who have created or are creating automatic trading agents. However they generally do not share any information about their techniques, or often even let on that they are using agents at all. Therefore, there may be significant innovations that are hidden from the public domain. Research competitions provide the incentive for people to develop similar innovations, and give us the ability to study their properties in an open forum.

Encourage flexible software and hardware. Taking a system out of one's own lab and into a new setting, whether it be a software system that is to be run on different computers or a robotic system that is to be run under different environmental conditions, requires a certain degree of flexibility in the system's creation. For example, rather than creating a vision system that works only in the lighting conditions in one's own lab, researchers must create a system that is easily adaptable to new conditions. Thus, the competition encourages general solutions that are more likely to apply in a wide variety of circumstances. In addition, since it is expected that rules of the competition may change slightly from year to year, it is always beneficial to create software that can be easily adapted to these changes.

It has been our experience so far that the benefits of RoboCup and TAC competitions outweigh the hazards. Most significantly as a strong source of inspiration, these competitions have played an important role in my own research [18, 21, 17]. Numerous other participants from both competitions have also published articles based on research originally motivated by RoboCup and TAC competitions (e.g [5, 14, 8, 10]). Again, the competition results themselves are not scientifically conclusive. But the process of competition, including the lessons learned, can be scientifically valuable.

6 Conclusion

RoboCup and TAC are both the focal points for large and growing research communities. The competitions play a large role in concentrating peoples' energies around consistent, challenging problems as well as providing them with concrete deadlines for producing complete working systems. Although there are significant potential pitfalls that need to be avoided when trying to facilitate research via competitions, the evidence from both RoboCup and TAC is that the benefits outweigh the hazards. From all indications, RoboCup, TAC, and perhaps other competitions that they will inspire will continue to play important roles within the research community in the foreseeable future.

Acknowledgments

Thanks to my collaborators on entries in the RoboCup and TAC competitions including Sorin Achim, Michael Bowling, János Csirik, Kwun Han, Michael Kearns, Michael Littman, David McAllester, Patrick Riley, Robert Schapire, Satinder Singh, and Manuela Veloso. Also thanks to the countless organizers of these competitions over the years.

References

1. Ronald C. Arkin. The 1997 AAAI mobile robot competition and exhibition. *AI Magazine*, 19(3):13–17, 1998.
2. Minoru Asada and Hiroaki Kitano, editors. *RoboCup-98: Robot Soccer World Cup II*. Lecture Notes in Artificial Intelligence 1604. Springer Verlag, Berlin, 1999.
3. Robert Axelrod. *The Evolution of Cooperation*. Basic Books, 1984.
4. Darse Billings. The first international RoShamBo programming competition. *International computer Games Association Journal*, 23(1):42–50.
5. Simon Ch'ng and Lin Padgham. From roles to teamwork: a framework and architecture. *Applied Artificial Intelligence*, 12:211–232, 1998.
6. Klaus Dorer. The magmaFreiburg soccer team. In M. Veloso, E. Pagello, and H. Kitano, editors, *RoboCup-99: Robot Soccer World Cup III*, pages 600–603. Springer Verlag, Berlin, 2000.
7. Anne Eisenberg. In online auctions of the future, it'll be bot vs. bot vs. bot. *The New York Times*, 2000. August 17th.
8. Nicoletta Fornara and Luca Maria Gambardella. An autonomous bidding agent for simultaneous auctions. In *Fifth International Workshop on Cooperative Information Agents*, number 2182 in Lecture Notes on Artificial Intelligence, pages 130–141, 2001.
9. Clemens Fritschi and Klaus Dorer. Agent-oriented software engineering for successful TAC participation. In *First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, Bologna, 2002.
10. Amy Greenwald and Justing Boyan. Bidding algorithms for simultaneous auctions. In *Proceedings of Third ACM Conference on E-Commerce*, pages 115–124, Tampa, FL, 2001.
11. Hiroaki Kitano, editor. *RoboCup-97: Robot Soccer World Cup I*. Springer Verlag, Berlin, 1998.
12. Hiroaki Kitano, Milind Tambe, Peter Stone, Manuela Veloso, Silvia Coradeschi, Eiichi Osawa, Hitoshi Matsubara, Itsuki Noda, and Minoru Asada. The RoboCup synthetic agent challenge 97. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*, pages 24–29, San Francisco, CA, 1997. Morgan Kaufmann.
13. Hugh Loebner. Home page of the Loebner prize. At `http://www.loebner.net/Prizef/loebner-prize.html`.

14. Stacy Marsella, Milind Tambe, Jafar Adibi, Yaser Al-Onaizan, Gal A. Kaminka, and Ion Muslea. Experiences acquired in the design of RoboCup teams: a comparison of two fielded teams. *Autonomous Agents and Multi-Agent Systems*, 4(2):115–129, June 2001.
15. Drew McDermott. The 1998 AI planning systems competition. *AI Magazine*, 21(2):35–55, 2000.
16. J. Rust, J. Miller, and R. Palmer. Behavior of trading automata in a computerized double auction market. Addison-Wesley, Redwood City, CA, 1992.
17. Robert E. Schapire, Peter Stone, David McAllester, Michael L. Littman, and János A. Csirik. Modeling auction price uncertainty using boosting-based conditional density estimation. In *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002. To appear. Available at `http://www.cs.cmu.edu/~pstone/papers.html`.
18. Peter Stone. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer*. MIT Press, 2000.
19. Peter Stone, Tucker Balch, and Gerhard Kraetszchmar, editors. *RoboCup-2000: Robot Soccer World Cup IV*. Lecture Notes in Artificial Intelligence 2019. Springer Verlag, Berlin, 2001.
20. Peter Stone and Amy Greenwald. The first international trading agent competition: Autonomous bidding agents. *Electronic Commerce Research*, 2002. To appear.
21. Peter Stone, Michael L. Littman, Satinder Singh, and Michael Kearns. ATTac-2000: An adaptive autonomous bidding agent. *Journal of Artificial Intelligence Research*, 15:189–206, June 2001.
22. Peter Stone, Patrick Riley, and Manuela Veloso. The CMUnited-99 champion simulator team. In M. Veloso, E. Pagello, and H. Kitano, editors, *RoboCup-99: Robot Soccer World Cup III*, pages 35–48. Springer Verlag, Berlin, 2000.
23. Manuela Veloso, Enrico Pagello, and Hiroaki Kitano, editors. *RoboCup-99: Robot Soccer World Cup III*. Springer Verlag, Berlin, 2000.
24. Manuela Veloso, William Uther, Masahiro Fujita, Minoru Asada, and Hiroaki Kitano. Playing soccer with legged robots. In *Proceedings of IROS-98, Intelligent Robots and Systems Conference*, Victoria, Canada, October 1998.
25. Michael Wellman. Personal correspondence, March 1999.
26. Michael P. Wellman, Peter R. Wurman, Kevin O'Malley, Roshan Bangera, Shou-de Lin, Daniel Reeves, and William E. Walsh. A trading agent competition. *IEEE Internet Computing*, 5(2):43–51, March/April 2001.