Dynamic Coalition Formation in Robotic Soccer

John Anderson and **Brian Tanner**[*] and **Jacky Baltes**
Autonomous Agents Laboratory
Department of Computer Science
University of Manitoba
Winnipeg, Manitoba, R3T 2N2, Canada
andersj,jacky@cs.umanitoba.ca, btanner@cs.ualberta.ca

Abstract

The ability to form coalitions of agents is central to multi-agent problem-solving. However, most multi-agent systems research still takes the view that teams are simply provided - an invalid assumption in most real-world situations. This paper describes an approach to forming coalitions of agents in robotic soccer, a domain where the dynamic nature of the environment plays a key role. We describe how agents that can learn about the abilities of others can form a coalition of the better-playing agents on the team, and show that this can be used to improve the performance of a team consisting of agents with varying skill levels. We also show that this mechanism is a useful one in a setting where agents are learning to play soccer, in order to form a coalition of agents from which to learn.

Introduction

The advantages of using of multiple agents to solve problems, both through basic division of labour as well the ability to economize on the cost of developing agents by dividing up specialized skills, is well known. Many approaches have been developed for decomposing problems, allocating tasks within a group, and combining results, using expertise from a wide array of fields, from game theory to sociology. Relative to this body of work, however, the study of the *formation* of teams and other types of coalitions is relatively immature. Most multi-agent systems research assumes existing teams with individual and group motivations that are reasonably well-understood, and work toward providing mechanisms to improve performance or ability – a few examples being (Fontan & Mataric 1996; Balch 1999; Anderson, Tanner, & Wegner 2002; Brogan & Hodgins 1997; Veloso & Stone 1998).

While such research is useful in improving our understanding of coordination and other important mechanisms, it does little to help us understand how such groups arise in the first place. What makes it worthwhile and useful for an individual to choose to belong to a group or not? Conversely, what makes it useful for a group to accept particular individuals and form a more functional team, while excluding others? In the real world not every individual who wants to be in a group is necessarily useful to it, and in order to have functional teams we may have to exclude agents from a group or work around the presence of less useful individuals.

Consider, for example, a human soccer team. While a professional team would be selecting their players from a small, select pool, teams below this level would more likely use a group of individuals with widely varying talents. As a player on that team, one would have to get to know the behaviour of one's teammates, and begin to form sub-groups of those with which it would be most useful to interact. Similarly, in an online community such as an e-commerce or distance education environment, there will be agents that have distinct advantages (e.g. more honest, more helpful, more knowledgeable) to most other members of the group than others. Loose coalitions of agents who believe they can be useful to one another form and dynamically change: the temporal extent of these may be momentary ("*I'll pass this to one of the good players*") or more long-lasting ("*I will never do business with a company that has one of my friends down*").

While any human belongs to many formal groups, these kinds of dynamic, informal coalitions - outside of or as part of more formal groups - are ubiquitous in human activity. We constantly make decisions in real time about who to interact with when we have a choice, and we reasonably quickly learn who in any formal group is there for the core purpose of that group, and who is there for other reasons (status as opposed to utility, for example). Most current work in coalition formation, however, focusses on formally defined groups – for example, (Tohme & Sandholm 1999; Yamamoto & Sycara 2001; Brooks & Durfee 2002) – as opposed to these more common informal organizations.

Our interest in these more informal organizations arises out of work in real-time decision-making in robotic soccer. Robotic soccer is a highly complex domain that has become a significant challenge problem in both mobile robotics and artificial intelligence. This challenge problem has been a tremendously useful to researchers in mobile robotics in terms of providing a common grounding for research in vision, control, and individual and team behaviour, among many other areas. Moreover, the use of a standard domain also affords the opportunity of competition in order to judge

[*]Current Address: Department of Computing Science, University of Alberta

the suitability of various proposed solutions in a controlled physical setting.

While it is certainly possible to have elaborate plays during a soccer game, the selection of these requires strong agreement beforehand simply because of the extremely limited time available in an ongoing soccer game to make selections and map roles through communication. We believe that the extreme time-limited opportunities that characterize robotic soccer make the formation of temporally short coalitions on a team both more interesting from a multi-agent systems standpoint and potentially more useful to exploit in the game itself. By such a coalition, we mean an understanding on the part of two or more agents that some opportunity can be exploited between them, due to recognition of their particular skills. For example, one agent might possess the ball, see an open teammate, receive a signal from that teammate that they understand the opportunity to achieve a shared goal, and choose to pass to that teammate based on some knowledge of that teammate's abilities. This involves some representation of the coalition on the part of the passing agent: the teammate being passed to has the necessary skills to belong to the group of agents deemed worthy of passing to under the game conditions currently being experienced. A relationship between the two players is momentarily formed over and above any other relationships that currently exist. This relationship may end immediately (the agent passes the ball and move on), or may continue for some time (the agent continues up an open channel in the field as the teammate works to keep that channel open). This is as opposed to an elaborate play relying upon understood relationships beforehand and pre-mapped actions as the play unfolds, or on individual behaviours that recognize no relationships with teammates (e.g. kicking the ball forward and a teammate simply happening to be there and moving it further, with no recognized connection on the part of either agent).

From the standpoint of knowledge of the coalition, each agent may have their own ideas of who constitutes members of the coalition (in this case, whose skills are worthy of being passed to), as opposed to a negotiated global membership list. Such a coalition as that described above is dynamic not only in that it is used to form momentary relationships, but also in the agents that constitute it. If someone seems to be having a bad day, for example, they may be excluded from an agent's concept of this coalition, and be placed back on it in future. New players may similarly get initial opportunities to prove themselves (an assumption of a skill level), and face a long period of earning trust if they fail in this initial period. While elaborate plays certainly exist in soccer and most other human sports, these lower-level decisions, and the changing coalitions of players we consider, form the backbone of moment-by-moment activity.

We have competed previously in the F-180 league at RoboCup (Anderson *et al.* 2003), and have employed this domain for research into issues in vision, teamwork, path-planning, and learning. We are working on developing agents that form informal, changing coalitions in this manner, adding and removing agents as they meet or fail to meet the criteria for that coalition. Such coalitions exist independently in each agent - there is no global membership list, and indeed, each agent's standards for membership may be different - however, the coalition itself is inexpensive to maintain in a real-time setting relative to the development of formal contracts, and we believe that such coalitions can greatly enhance the performance of teams in many real-time settings. Our intent is to empirically examine the use of such informal coalitions over teams that do not employ such structures.

This paper describes recent work toward this end. In order to judge the behaviour of an agent over time, and decide whether it should be included in a coalition, we must be able to learn about the behaviour of others, and so we begin by describing relevant work in multi-agent reinforcement learning that supports this. We then describe our approach to forming coalitions and the implementation of the agents involved. Finally, we compare the performance of agents using various elements of these coalition formation techniques to teams of agents with varying abilities that do not take this approach, and illustrate that this is useful in learning from other agents as well as simply performing better as a team.

Learning about others as a basis for coalition formation

In order to be able to form coalitions from useful agents, we need to determine how useful particular members or potential members of the coalition are to us. To do this, we must either use pre-supplied information (first impressions, prior records), and/or have the ability to learn about the behaviour of others over time. We are most interested in employing this in domains where a group of agents are placed together with no prior knowledge, so we currently employ no initial opinions about other agents: any agent has no particular reason to believe anyone else may perform well or poorly. We can begin with any initial opinion desired however, allowing for inclusion of prior reputations.

In order to be able to form coalitions of agents that are useful for various purposes, an agent must be able to learn about the abilities of other agents. In soccer, we learn about other players through observation during play on the field. We can supplement these with coarser statistics (e.g. total goals a player has scored), but we can more quickly learn simply by watching for episodes of good and bad behaviour as play unfolds (since there may be many episodes of either during the time leading up to a goal being scored). There are any number of indicators of a good or bad player: passing a ball toward one's own goal rather than away, where either option is available, is not particularly strategically useful, for example. This can be extended to use the models we are maintaining as agents: if I know that player X is a poor player, and I see you pass the ball to player X at a crucial moment when other players could have been employed, this might lower my opinion of you. One of the difficulties in any multi-agent situation is the question of individual perspective: the situation may look different than it actually is because of the angle of view, personal knowledge (I may not know some action is as bad as it turns out to be), visual occlusion, etc. Like other learning settings, we have to deal with the fact that we will get inaccurate perspectives sometimes, and ensure that (all things being equal) the effect of repeated accurate experiences will outweigh the few incorrect ones.

As activity in the domain unfolds, agents get a more and more accurate picture of others and form coalitions that allow better results to be achieved than would be possible without such knowledge. The degree to which behaviour is improved is, of course, directly affected by the variation in skill levels between agents: if everyone is very good in all ways, for example, devoting any time to choosing who to interact with on the basis of skill will not further improve team behaviour. However, the greater the number of poorly-skilled (or to generalize to other domains, dishonest or otherwise untrustworthy) agents, the greater the difference in performance the use of models of others as a basis for informal dynamic coalitions should be. The domain also affects the utility of this: the greater impact any poorly-skilled agent can have, the more important it is that we have a mechanism to avoid such interactions.

There is also the issue of maintaining membership in any agent's view of the coalition over time. Agents associate with one another for many reasons, and it may be that whatever skills an agent brought to a coalition may be lost, that an agent may have been deceptive about its abilities when it became a member, or that that the agent's own priorities may have changed over time and it may no longer be keeping up with any services to the coalition. Whatever the reason, agents must be forced out of such informal coalitions when they no longer useful. In the case of a game like soccer, this is an interesting situation. If I observe a wide range of skills in teammates, and I wish to work with only the best of these given the choice, I can learn about others over time and gradually form a coalition of useful members. If one of those agents' skills drastically change (e.g. an injury, or in a robotic context a power failure) , I want to immediately note that and remove them from the coalition. If I do this to harshly, however, it's possible for an agent making only a few momentary mistakes to be excluded completely where that agent in the long run would be of benefit to the coalition. In a real robotic domain especially, power and other types of failures (e.g. being lost from the view of an overhead vision server) are common and often short-term. We need a method that will both exclude poor agents and yet be flexible enough to deal with short-term failures. These aspects of our approach will be described in Section , following a discussion of the manner in which we have implemented these agents.

In addition to maintaining coalitions of agents for the purposes of enhancing play on a mixed-skills soccer team, there are also other intriguing possibilities for the use of dynamic coalitions. If I want players to be able to improve, I can either use an external teacher such as a coach, or I can further attempt to take advantage of the presence of multiple agents to allow agents to learn from each other. We have been experimenting with forms of reinforcement learning during soccer play that allow a much greater quantity of reinforcement to be gathered from the observations of other players during play, as opposed to strictly external sources such as goals scored, or sources that are only intermittently available, such as a coach during time-outs (Anderson, Tanner, & Wegner 2002). The difficulty with this in practice is that while the number of potential reinforcements is large, the quality of many of these is poor. Given that reinforcement is based on observation, an agent that plays soccer well will provide useful reinforcement in general (subject to the same inaccuracies due to individual perspective as those detailed above). An agent that is poorer will likely give information that is not useful or even counterproductive to learning to play good soccer. Learning in such a setting can only progress by developing the ability to distinguish which agents are worth learning from (or, considering the problem from a finer grain, the *degree* to which each agent should be considered a good model of behaviour in the domain). From the standpoint of the dynamic coalitions discussed above, we must develop coalitions of agents to learn from, and adjust these based on perception of those agents' abilities.

The ability to form a coalition of agents to learn from can be supplied by the same coalition-formation mechanism used for any other informal coalition. However, in order to allow learning of good soccer player behaviour, we must add an additional learning mechanism to the agent. We employ reinforcement learning techniques to learn soccer (Anderson, Tanner, & Wegner 2002), and our approach to reinforcement learning uses a situation-action mapping for soccer-playing behaviours represented in the form of a basic table (as has been done in robotic domains in many prior robotic learning implementations, such as (Mataric 1997a; 1997b)). A variant of Q-learning (Sutton & Barto 1998) is employed to take reinforcement from others and spread it back through the sequence of actions that have unfolded. This mapping forms the agents' current set of behaviours, and improves over time. This same mapping, whether learned or not, also serves as the basis from which the behaviour of other agents is judged: agents employ their own abilities to ask what they would do when they see the behaviour of others, and use this to update their model of the abilities of other agents.

The next section describes the implementation level of these agents. Following that, we show the utility of employing such informal coalitions to improve the performance of a robotic soccer team without having the players learn from another (Section . Then (Section , we show that coalitions can be used to assist the peer reinforcement learning mechanism described above to learn from whom to take reinforcement.

Implementation

Our implementation of this approach is done using the RoboCup Soccer Server (Noda *et al.* 1998) version 8 for Linux. Agents are behaviour-based and written in Java, and the server simulation is run at 4x speed (1 cycle=25ms) in order to allow learning to occur relatively quickly. Each agent has basic behaviours for ball handling, passing to other players, kicking on goal, and offensive and defensive movement. The weighting of each of these behaviours is varied based on the agent's own perceptions of how offensive or defensive the setting is (e.g. standing in front of an opponent's goal vs. one of the opposing team's players coming with the ball into our own end). Agents maintain a very simple world model, storing only the location of the ball and other players on the field (relative to flags placed on the field by the server). To reflect the dynamic nature of the game, this model decays very quickly over time - complete lack of confidence in the position of any object is reached in less than a second. Agents do not perform any tracking over time of objects (i.e. there is no information tracked over more than 1 visual frame, such as the ball travelling from one player to another), because the speed of the game goes against having such elaborate world models. Instead, relative positions of useful objects are noted (for example, a player being at roughly the same location as the ball, or behind the ball and in turn behind the goal for a kick), and events such as likely kicks, passes, and goals scored are inferred from these. This leads to perception that is fallible as it would be in the real world.

In order to provide players of varying skills, we have two options. We can either handicap player's ability to perceive the world (if I can see the ball further than some other player, I have a distinct advantage), or by altering their behaviours (if I am less accurate in my shot placement, or can kick the ball only a much shorter distance, for example, I am decidedly disadvantaged). Although either of these allow variation in player skill, it is more realistic to employ the latter, and we thus handicap some players by giving them less accurate ball handling abilities - they are more likely to miss a target or have a shot be short or long. We could also have all players learn all skills from scratch, which would naturally have everyone be extremely poor at the start. We have thus far not attempted this, in part because of the length of a trial, but also because of a desire to control performance enhancement from learning basic skills vs. performance enhancement from learning to interact in a coalition.

In terms of maintaining reputations, the soccer domain is different from many other multi-agent domains in that we are not interested in agents that are deceptive but simply those that are not likely to perform well. To maintain information about other players we employ a pragmatic, heuristic approach that is manageable in real time rather than attempting to use elaborate statistics. As stated above, learning about other players is done purely through observation. Each player maintains an ongoing reputation in terms of a cumulative average of episodes observed for each player, and notes the following types of episodes to form a model of the skills of other players: observing a player gaining possession of the ball in the open enhances the reputation of that player (+4), as does gaining possession from an opponent (+4) and scoring goals (+10). Losing possession to an opponent lowers reputation (-4). The latter will happen much more often in players that are poorly skilled, since they will not be able to control the ball as well when it is kicked. Because of the heuristic perception employed, we do not have a constant stream of updates: on average, a player views one of these significant actions on the part of a teammate approximately every 7.5 seconds.

The teams that we employ initially know nothing about one another, and consequently the reputations maintained by any player fluctuate wildly initially. These settle down in the long term, and when agents are static (i.e. they are not learning new behaviours) their reputations settle into relative consistency after approximately 1.5 hours of play (200,000 time steps). Figure 1 shows the evolution of the reputation of a typical good and bad agent over time. Reputations will always fluctuate slightly since it is possible for agents to see several good things or bad things in a row, and none of the agents is perfect.

Effect of Coalition Formation on Team Play

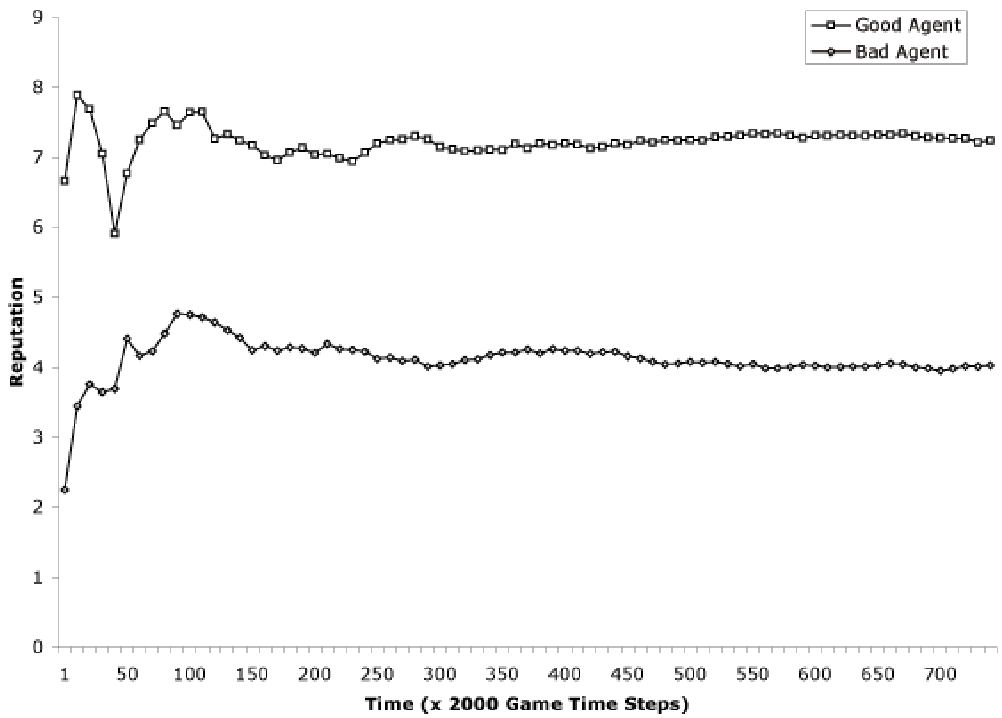
In order to decide whom to work with, we need some basis for distinguishing agents that are worthwhile from those that are not. Since we are dealing with teams that know nothing about one another initially, we have no absolute means for determining a boundary between a good and bad agent. Instead, we employ the following heuristic methodology.

With reputations based on observation, an agent observes many more episodes of its own behaviour over time than it does of others. This is reflected in a more accurate view of the agent's own abilities than those of others. If we assume we are like most other agents, we can use the difference between how we see others and ourselves as an estimate of error - an agent that is like ourselves could look as different as this or more to us. Since we have a range of skill levels, however, we can't assume we're like all other agents. In order to get a better estimate of error, we allow agents to inquire about their own reputation from others, allowing a more direct direct measure of the potential error between our own perception and how others see us. This does not assume others are like us in soccer skills, only that they employ the same mechanism for maintaining information about others. This is not basing reputations on communication - we never share the reputations of other agents, we ask only about ourselves (although we have found that sharing all models of other agents using communication does not speed up the accurate modelling of others over observation as much as would be expected). We use a multiple of this difference as a form of standard deviation that can be used to reliably distinguish good agents from bad. This is easily employed in real-time, since our models of others stabilize after a reasonably short time and we do not have to continually inquire about how others see us.

While we have found that this is a useful method for judging bad agents where the effect of a bad agent on team play is linear or super-linear (that is, the effect on the score of a game when a good agent is substituted with a bad one decreases linearly or worse than linearly), it is not as useful when bad agents do not have as negative an effect (i.e. they are sub-linear in terms of their effect on score). We are currently working on a less relative approach by adapting Balch's concept of *social entropy* (Balch 1998) to deal with measuring the effect of a bad agent on a team's skills (i.e. viewing poor skills as a form of heterogeneity) and attempting to quantify the direct effect of a bad agent.

The effect of using coalitions to decide with whom to interact with depends, of course, on precisely what we use those coalitions for. If we have formed a collection of what

Figure 1: Developing reputations over time



we consider "good" agents, we can choose to use it to selectively interact with other players - for example, to choose to pass the ball to good agents over bad ones when there is some option. We can also employ a coalition more conservatively, by actually working to exclude agents that are not part of the coalition from play - for example, by moving to intercept a pass to a non-coalition agent. There is, of course, a range of behaviours in between as well. Figure 2 shows the effect of using this mechanism in both of the above situations individually and in combination, on a team where the presence of a bad agent is super-linear. The figure shows the score ratio of a team employing this mechanism to form a coalition of good agents compared to an otherwise-identical team that does not employ this coalition mechanism.

Maintaining Coalitions

Being part of a coalition also should mean that membership has some meaning to an agent - that it is economically (or in some other sense) desirable to maintain membership. Since these coalitions exist within the perspective of a single agent, this translates to it being undesirable to remove someone from the coalition for observing infractions that do not undermine the economic contribution the offending agent's participation makes. If an agent witnesses repeated severe transgressions (e.g. scoring repeatedly on one's own goal), it is likely worthwhile considering the offending agent bad no matter what its prior reputation. Temporary failures, however, happen often in robotics and must be dealt with without long-term harm to a reputation in order to allow the coalition to reap the benefit of the agent's participation once it is again available.

We have been exploring issues in the length of memory and its effect on agent reputation in order to make coalitions of the type described in previous sections more receptive to change on the part of the abilities of the agents being modelled. In particular, we are interested in modelling failures in agents and adapting the reputation modelling to these, since temporary agent failures are common in robotic soccer. Purely long-term reputations will not suffice for this purpose, because it takes too long for a reputation to degrade enough for the agent to be recognized as bad and thus excluded from the coalition (i.e. we keep passing the ball to an agent that is dead on the field), and then too long once that agent gets started again to eventually be included. Adding a short-term reputation mechanism for such settings has proved useful. The short-term memory consists of a reputation based on the last *N* observations of that agent rather than the long-term running average. The effect of an N-step short-term memory (N=10,30,80) on the reputation of a failing agent is shown in Figure 3. This shows the points where an agent was disabled and then re-enabled, and the delay in reputation decay (and later improvement) by memory length. Of the various lengths we have been experimenting with, we have found only a very short (10-step) short-term memory to be effective in dealing with properly managing reputations for failures of the lengths we have anticipated for a robotic soccer domain (10 minutes). In terms of the effectiveness of employing this on a soccer team, Figure 4 illus-

Figure 2: Effect of employing a coalition to choose an agent in a passing situation, in moving to intercept the ball, and both in combination

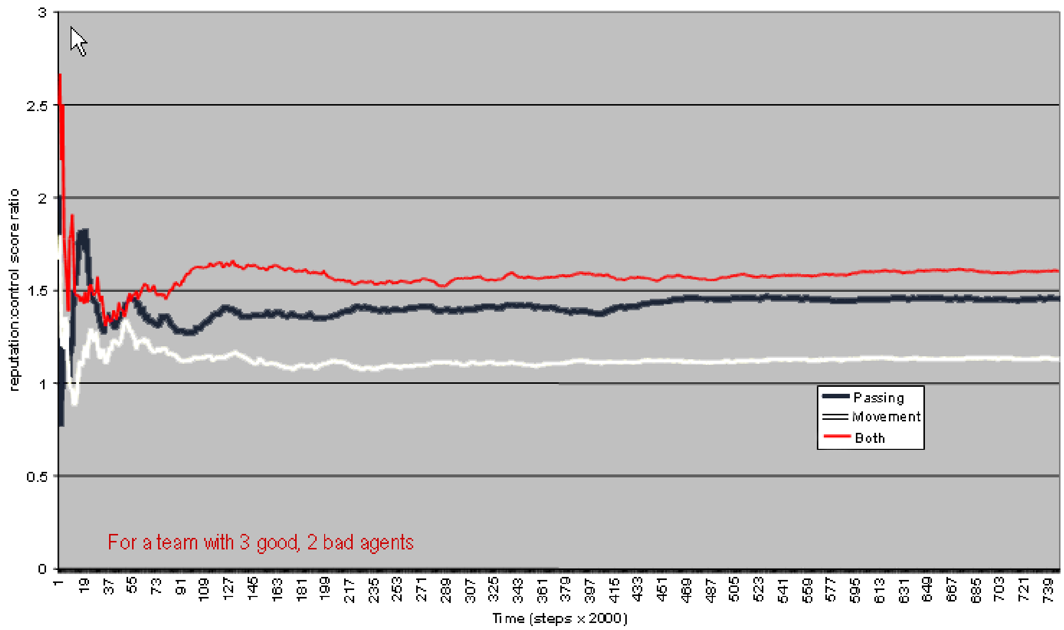


Figure 3: Reputation decay for various lengths of short term memory

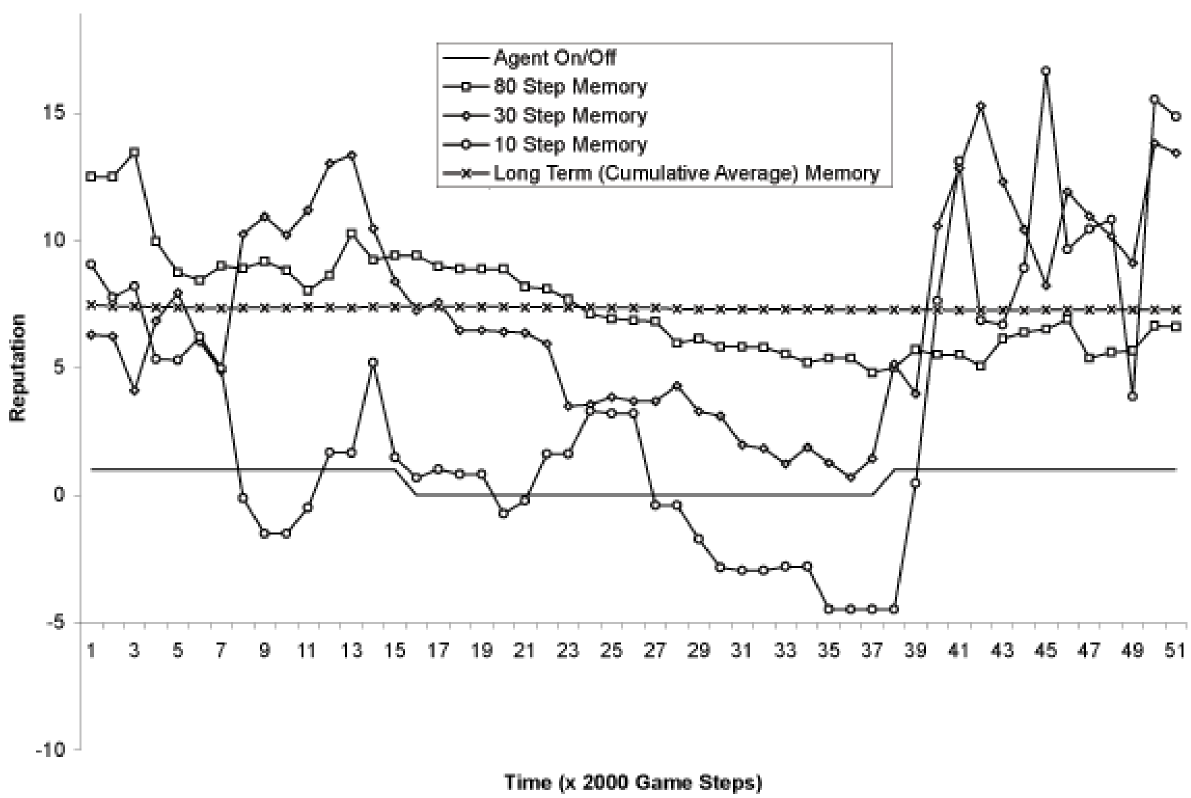
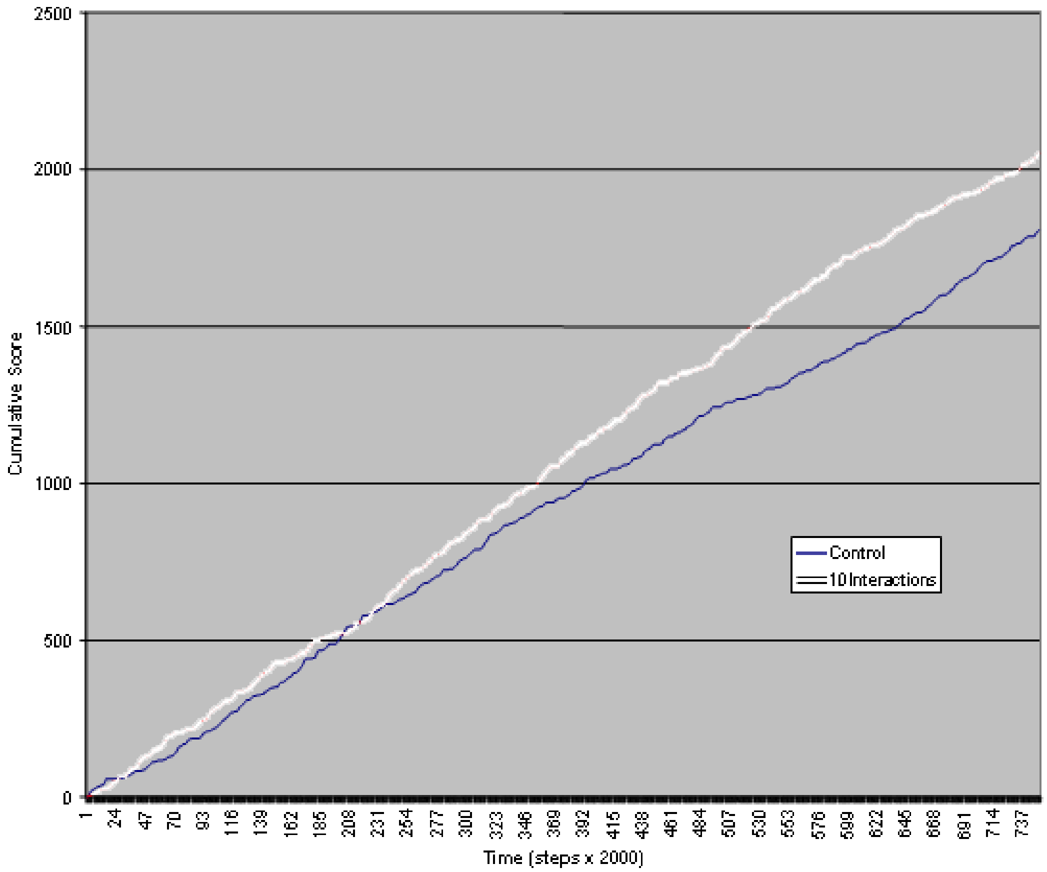


Figure 4: Effectiveness of short- and long- term memory alone where agents can fail



trates a comparison between a control team (using regular, long-term reputations as above) and a team supplementing this with the 10-step short-term memory, under conditions where the likelihood of any agent failing on a particular time cycle is 0.00025, with all failures lasting ten minutes. This translates to an average of one failure every fifteen minutes of play. The agents themselves have no concept of what a failure is, they know only (from observation over time, using the modelling mechanisms described earlier) that an agent is not playing the way a good agent would be expected to. The actual calibre of play of the agents is exactly the same as the experiments shown earlier. Short memories, as would be expected, are not useful on their own. The relative comparison of teams forming coalitions using reputations based *purely* on short- and long-term memory, in comparison with control teams using each type of memory and not employing coalitions, is shown in Figure 5. This shows that while short-term memories are useful for dealing with temporary setbacks in reputation (such as these minor failures), they are not as useful on their own as a long-term reputation mechanism for forming coalitions of good agents. We are continuing to explore combining short- and long-term memories in a dynamic fashion to explore issues of balancing the likelihood of exclusion from a coalition of good agents with the potential economic benefit of the agent.

Using Coalitions to Improve Learning

We began this work not just so that we could improve the behaviour of teams where agent skill is diverse through the formation of coalitions, but so that we could use reinforcement learning where reinforcement came from peer agents, gradually learning who to learn from. This can be considered forming a coalition of teachers from the learner's viewpoint.

To examine the effectiveness of this, we first determined a simple collection of 9 correct-perception action mappings that would translate to better performance on a soccer team (for example, kick when you are close to the goal and have the ball; move the ball toward the opponents goal). These behaviours were to be learned, while the many other behaviours necessary to play soccer were provided to all agents. We used the methods of reinforcement described in Section : each agent observes other agents and issuing positive or negative reinforcement based on what that agent would itself do in the same situation, and a learning agent takes reinforcement from other agents, considering it in light of the agent's model of the source(s), and alters its own perception-action mapping accordingly.

To examine the efficacy of the coalition-formation mechanisms on this process, we took an agent learning by reinforcement from others, and examined the effect of modelling other agents's abilities described in Sections and on the amount of learning that took place and the amount of time this took. A single learning agent was placed on a team where all other agents were not learning soccer-playing behaviour. The learning agent was given a subset of the abilities that a good soccer playing agent would require. Basic movement was complete, but a subset of behaviours involving decisions as to what to do when the agent had the ball in any particular state (shoot it at the goal; pass forward to a teammate; pass backward to a teammate; or dribble the ball towards the goal) were learned. In order to have an agent that had at least some ability to do something with the ball initially, some of the poorest choices were pre-filtered: for example, when an opponent was within 5 meters of the agent, dribbling the ball to the goal was not provided as a potentially viable option. Considering the situations in when each choice could be active and possibly observed by a teammate in order to get reinforcement, this amounted to 9 particular states for which the agent had to learn the action that would be considered correct for a good soccer-playing agent.

In this situation, the amount that could be learned was directly affected not only by the degree to which the agent could differentiate accurate from inaccurate reinforcement, but by how much reinforcement was actually accurate - that is, how many of the agents on the team were actually good players. To control for this we set up two team categories: a team where the learning agent had four equally good teammates to learn from, and another where the team was equally split by having two good teammates and two with poor skills. While those with poor skills would still be of some benefit in learning (that is, to the degree their skills were at all useful), presumably the agent would learn as almost much bad behaviour as good without the ability to discern reinforcement from agents that actually could perform well from reinforcement from poorly-performing agents.

Two instances of each team were created, one a control pair that used peer-based reinforcement learning with no attempt to form a coalition of good agents from which to learn, and a second pair that formed a coalition in order to ignore reinforcement from agents that were deemed ineffective. Figure 6 shows the results of this learning over time, in terms of the cumulative average of the states correctly learned over time. Figure 6 shows that forming a coalition has no effect over not forming a coalition in the case where all agents are good - this is natural, since the best one can achieve with the coalition is to have all agents involved, which is already done in the control group. Moreover, when employing coalition formation in a group consisting of entirely good agents, the coalition formation mechanism actually makes the learning process less successful at first. This is because useful reinforcement is not taken seriously until the agent realizes that its teammates are in fact good agents whose reinforcement is worth using. On the other hand, if 50% of the teammates are poorly skilled, forming a coalition to learn from the good agents is enormously beneficial. In fact, a coalition of 2 good agents was almost as useful as having only skilled teammates once the reputations other agents were learned and the coalition employed. The difference in successful learning between this group and a non-coalition forming team with four good agents is due both to the partially inaccurate reinforcement received from the two poorly-performing agents, as well as the accurate reinforcement that was discounted during the time reputations were being formed.

Discussion and future work

We have seen that the ability to dynamically form coalitions is useful in robotic soccer, both in terms of the effectiveness

Figure 5: Comparison between using only short-term and long-term memory for agent models

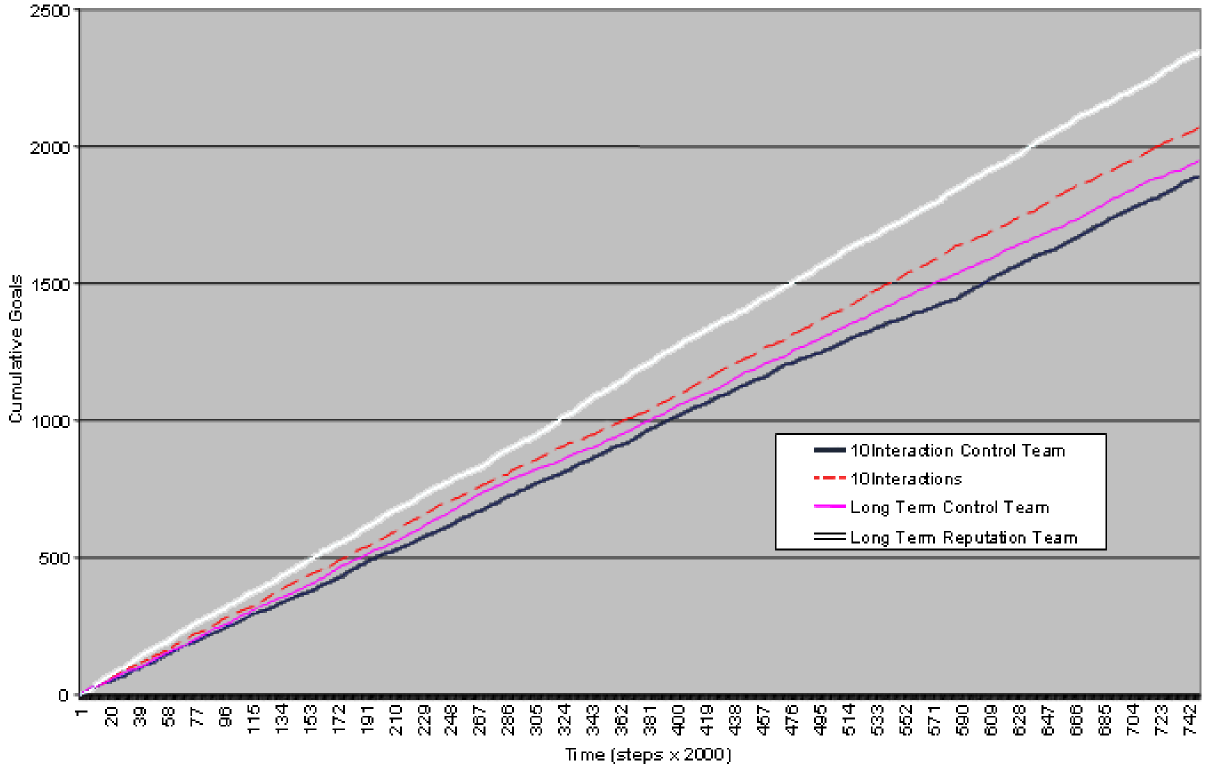
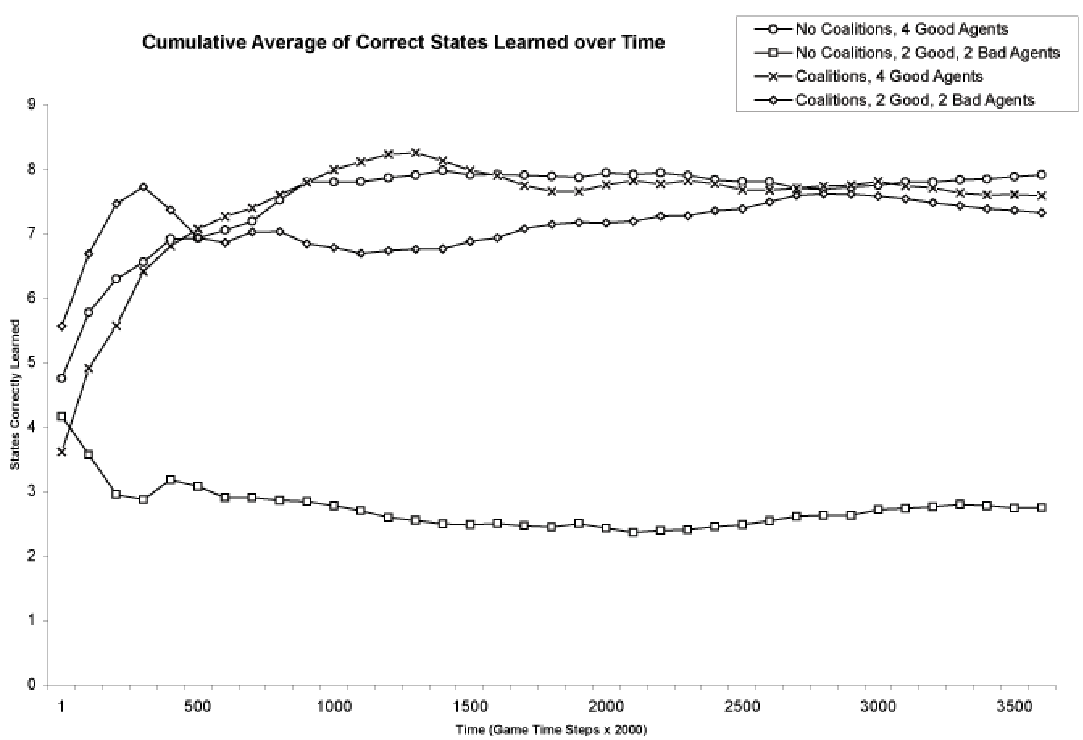


Figure 6: Learning: employing vs. not employing coalitions



of limiting the damage that less skilled players can do, and in terms of selecting players to learn from in a learning situation. We have also shown that the exclusion of players from a coalition is assisted through the use of a short-term memory as opposed to purely relying on long-term reputation.

In addition to performing further experimentation on the nature of dynamic coalitions in this domain and its effect on group reinforcement learning, there are a number of intriguing areas for further work. If one's reputation is the key to how one is treated by others, for example, behaviour to safeguard reputations should be important. Here there is no global membership associated but agents know that others are using the same coalition management skills. We also want to use coalitions to allow agents to learn to refrain from committing actions that will damage their reputation in the eyes of others - including trusting others too easily, and the effect of this on the play of a soccer team. This will allow us to apply this to settings where deception, for example, is a possibility.

We are also interested in the real-world aspects of more formal forms of coalition formation, where membership is posted or shared globally (as opposed to these coalitions where each agent could possibly have a different concept of the coalition). There are many real-world elements that have yet to be considered properly by such mechanisms, such as agents that are using the coalition for their own purposes as opposed to that for which the coalition officially exists (van de Vijsel forthcoming). We see the most important future work on coalition formation to be on bridging the gap between these two mechanisms: moving from the informal coalitions formed in a moment and lasting for as long as the coalition is useful, to more formal elements where guidelines for behaviour are developed between the coalition's members. In the work presented here, for example, short-term reputation mechanisms are currently used to deal with potential agent failure. If an agent could propose more formal bonds between members that have a high reputation in one another's eyes, we could potentially do away with such mechanisms by allowing trusted members to remain trusted despite a failure (use such mechanisms to simply recognize an agent as unavailable as opposed to genuinely bad), and forcing members who have not yet earned a high reputation to earn back their reputation the hard way.

References

Anderson, J.; Baltes, J.; Cornelson, D.; Liu, T.; Stuart, C.; and Zilkie, A. 2003. The little black devils. In *Proceedings of the RoboCup Symposium*.

Anderson, J.; Tanner, B.; and Wegner, R. 2002. Peer reinforcement in homogeneous and heterogeneous multi-agent learning. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing (ASC2002)*.

Balch, T. 1998. *Behavioural Diversity in Learning Robot Teams*. Ph.D. Dissertation, Computer Science, Georgia Tech.

Balch, T. 1999. The impact of diversity on performance in multi-robot foraging. In *Proceedings of the Third Annual Conference on Autonomous Agents*, 92–99.

Brogan, D. C., and Hodgins, J. K. 1997. Group behaviours for systems with significant dynamics. *Autonomous Robots* 4:137–153.

Brooks, C. H., and Durfee, E. H. 2002. Congregating and market formation. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, 1–100.

Fontan, M. S., and Mataric, M. J. 1996. A study of territoriality: The role of critical mass in adaptive task division. In Maes, P.; Mataric, M.; Meyer, J.-A.; Pollack, J.; and Wilson, S. W., eds., *Proceedings, From Animals to Animats IV, Fourth International Conference on Simulation of Adaptive Behaviour*, 553–561. MIT Press/Bradford Books.

Mataric, M. J. 1997a. Learning social behavior. *Robotics and Autonomous Systems* 20:191–204.

Mataric, M. J. 1997b. Reinforcement learning in the multi-robot domain. *Autonomous Robots* 4(1):73–83.

Noda, I.; Matsubara, H.; Hiraki, K.; and Frank, I. 1998. Soccer server: a tool for research on multi-agent systems. *Applied Artificial Intelligence* 12(2-3):233–250.

Sutton, R., and Barto, A. 1998. *Reinforcement Learning*. Cambridge, MA: MIT Press.

Tohme, F., and Sandholm, T. 1999. Coalition formation processes with belief revision among bounded-rational self-interested agents. *Journal of Logic and Computation* 9:1–23.

van de Vijsel, M. forthcoming. Coalition formation in multi-agent systems under real-world conditions. Master's thesis, Department of Computer Science, University of Manitoba, Winnipeg, Canada.

Veloso, M., and Stone, P. 1998. Individual and collaborative behaviours in a team of homogeneous robotic soccer agents. In *Proceedings of the Third International Conference on Multiagent Systems*, 309–316.

Yamamoto, J., and Sycara, K. 2001. A stable and efficient buyer coalition formation scheme for e-marketplaces. In *Proceedings of the International Conference on Autonomous Agents*, 576–583.