The RoboCup 2014 SPL Drop-in Player Competition: Experiments in Teamwork without Pre-coordination

Katie Genter
Department of Computer Science
University of Texas at Austin
Austin, TX 78712 USA
katie@cs.utexas.edu

Tim Laue
Department of Computer Science
University of Bremen
28359 Bremen, Germany
tlaue@uni-bremen.de

Peter Stone
Department of Computer Science
University of Texas at Austin
Austin, TX 78712 USA
pstone@cs.utexas.edu

Abstract

The Standard Platform League is one of the competitions of the annual RoboCup world championships. In this competition, teams of five humanoid robots play soccer against each other. In 2014, the league added a new sub-competition which serves as a testbed for cooperation without pre-coordination: The Drop-in Player Competition. Instead of homogeneous robot teams that are each programmed by the same people and hence implicitly pre-coordinated, this competition features ad hoc teams, i. e. teams that consist of robots originating from different RoboCup teams and that are each running different software. In this paper, we provide an overview of this competition, including its motivation, current rules, and latest results.

1 Introduction

As robots become more prevalent in the world, they are increasingly being designed to work in teams to accomplish tasks. Usually, all of the robots on a team are programmed by one organization, and hence are implicitly designed to work together in a specific way. RoboCup, an annual international robotics competition, features many such teams that are programmed by universities, organizations, and companies to play soccer in various leagues. This paper presents a specific competition held in the Standard Platform League at RoboCup 2014 in which teams were encouraged to develop 'drop-in' soccer players that could be good teammates and play well within a team composed of drop-in players from a variety of teams in the Standard Platform League.

In the Drop-in Player Competition discussed in the paper, each team programmed a robot to coordinate with unknown teammates. The teams were asked not to pre-coordinate, so that during games these agents had to engage in *ad hoc teamwork* in order to reason about their teammates' abilities and intentions in real time and determine how to best assist their team. Each agent's goal was to win the soccer game by as much as possible, while being judged as a 'good teammate' by human officials that were watching the game.

It is often challenging when working with real robots to gather extensive experimental data. The 2014 Standard Platform League Drop-in Player Competition gathered robotic agents from 23 teams, involved at least 50 human participants, and consisted of fifteen 20-minute games for a total playing time of 300 minutes. With 10 robots participating in each game, this totals to an experiment utilizing 50 robot hours! Hence, this competition proved to be the largest ad hoc teamwork experiment on robots that the authors are aware of to date, and is likely one of the largest robotic experiments involving as many as 23 different organizations.

The 2014 Standard Platform League Drop-in Player Competition grew from a technical challenge held at RoboCup 2013 in three different leagues (MacAlpine et al. 2014). The technical challenge in 2013 was optional, and only saw six teams participate in the Standard Platform League. The authors of this paper helped plan, organize, and run the substantially larger Drop-in Player Competition at RoboCup 2014. This paper serves as a follow-up paper as it details the 2014 Drop-in Player Competition in the Standard Platform League and highlights the advancements in the competition as well as in the strategies utilized by various teams.

Section 2 starts by describing the Standard Platform League as a RoboCup league and introduces the concept of ad hoc teamwork. We provide details pertinent to the Drop-in Player Competition in Section 3. One of the major contributions of this paper is the description given in Section 4 of the strategies employed by various teams. Section 5 presents the results of the 2014 Drop-in Player Competition, including both the judged scores and the goal differential scores, and then analyzes these results. Section 6 reviews research related to the competition, and Section 7 concludes.

2 Background

Two important areas of background knowledge are introduced in this section. The first is the Standard Platform League of RoboCup and the second is the multi-agent systems research area of ad hoc teamwork.

2.1 RoboCup Standard Platform League

RoboCup is an international robotics competition that was founded and had its first competition in 1997. The first competition saw just 38 teams from 11 countries, but the competition has grown to include 410 teams from 45 different countries in 2013. Within RoboCup there are currently five major competition divisions, one of which is RoboCup Soccer. Within the RoboCup Soccer division, there are eight leagues including the Standard Platform League (SPL). The SPL is different from the other leagues in the RoboCup Soccer division in that all teams must use the same robotic platform, making it essentially a software competition.

The SPL was first held in 1999 as the ‘Four Legged League’. In 2008 the league transitioned from using Sony Aibos to Aldebaran NAOs and became known as the ‘Standard Platform League’. The league has grown from 15 teams using the NAOs in 2008 to 24 teams in 2014. Historically, teams have competed in a main soccer competition as well as various technical challenges. The main competition is usually executed as one or more round robins where the top teams from these round robin pools gain spots in an 8-team single-elimination bracket. The technical challenges are optional competitions lasting no more than two hours in which teams compete in announced challenge tasks that are designed to advance the league. A smaller version of the Drop-in Player Competition was held as a technical challenge in 2013 before becoming a separate SPL competition in 2014.

Teams in the SPL compete in 5 on 5 soccer games on a 9 meter by 6 meter soccer field. Each game consists of two 10-minute halves, where the clock stops after goals only in the semi-finals onward. Teams must play completely autonomously — no human input is allowed during games outside of game state signals sent by an official to communicate to the robots when a goal has been scored, when they have been penalized, and so on. The playing environment is color-coded — goals are yellow, lines are white, the field is green, and players wear either maroon or cyan jerseys. The robots on each team are allowed to communicate over a wireless network. See Figure 1 for a picture from an SPL game.

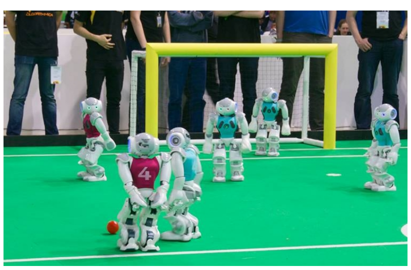

Figure 1: NAO robots playing in a SPL game.

2.2 Ad Hoc Teamwork

Since 1997, RoboCup has served as an excellent domain for testing teamwork, coordination, and cooperation. Most teams have successfully programmed their robots to work well as a team, coordinating which robot should go to the ball, which robot should play defense, and even what formation should be adopted by the team when facing various different opponent types. However, the 2013 drop-in player challenge across three RoboCup leagues (MacAlpine et al. 2014) was one of the first organized efforts to evaluate a player’s ability to coordinate with a set of teammates in an ad hoc manner, and the 2014 SPL Drop-in Player Competition greatly improved upon the 2013 challenge in both scale and participation.

Ad hoc teamwork is different from most research on teamwork because it focuses on creating agents that can cooperate with unknown teammates without prior coordination. Stone *et al.* imagined ‘staging field tests of ad hoc team agents at the annual RoboCup competitions’ in their 2010 AAAI challenge paper that introduced ad hoc teamwork (Stone et al. 2010). The SPL Drop-in Player Competition at RoboCup 2014 did just this. By organizing the SPL Drop-in Player Competition as a well-organized competition in which all SPL teams must participate, the authors and RoboCup trustees have created the potential for a longstanding empirical testbed for ad hoc teamwork research.

3 Competition Description

The Drop-in Player Competition is based on the normal RoboCup SPL soccer competition, i. e. it consists of soccer matches, in which teams of 5 robots play against each other. However, to make it a meaningful competition about teamwork without pre-coordination, several changes and additional preparations are necessary.

3.1 Altered Rules of the Game

Basically, the rules of the Drop-in Competition games are the same as for normal robot soccer games in the SPL. There only exists one major difference concerning role assignment. In normal SPL games, the player number *1* is the only player that is allowed to play as a goalkeeper, i. e. to permanently stay inside the own penalty area and to touch the ball with its hands while being in this particular area. As the robot numbers are assigned randomly (see 3.4), such a predefined role assignment is unwanted because it would assign player number 1 to be the goalkeeper. Instead, the robots on the field must arrange the role assignments for themselves. The first robot that enters his own penalty area will be considered as the goalkeeper for the remainder of the game. All other robots are automatically considered to be normal field players. Note that this may lead to games in which no goalkeeper exists, as its existence is not enforced externally.

3.2 Standard Communication

To enable communication among the players of a team, the SPL introduced a wireless standard communication protocol in 2014. This protocol is mandatory for all normal SPL games as well as for all SPL drop-in games. No other wireless communication is allowed. Technically, each robot is allowed to send up to five UDP broadcast messages per second to its team. Each message has a predefined format and includes information about the robot’s position, walk target, shooting target, the observed ball state, and the robot’s intention. The latter is a state that could be *want to be goalkeeper*, *want to play defense*, *want to play the ball*, *I am lost*, *nothing* (default). However, all communication is unidirectional and no negotiation mechanisms exists. Therefore, each robot can express its intention but there is no guarantee that its teammates actually consider this in their own decisions.

For normal SPL games, the standard message contains an additional data block that is used by teams for team-specific information and data structures. Technically, this block also exists in the Drop-In Competition, but is generally not considered to be useful since robots programmed by other teams do not know how to interpret this data.

3.3 Scoring Scheme

Each player's score in the Drop-in Player Competition consists - in contrast to the normal soccer competition - of two equally weighted components computed over all of the player's games: average goal difference and average judge score.

Average Goal Difference The aim of a soccer match is to score more goals than the opponent team. Although the Drop-in Player Competition is about teamwork without precoordination, it is set in a soccer scenario and the essence of this scenario should be preserved. Hence, over all games that a robot is scheduled to play in the competition, the average goal difference is computed. Therefore, the intentions of each robot should always be to score goals as well as to prevent the opponent from scoring. The competition does not involve any individual rewards for scoring goals to make both types of role – defensive and offensive – equally attractive within a team. However, each robot has to be aware of its own skills and thus of its options to best support its team.

Judge Scores Similar to human soccer, important aspects of good team play, such as passing and good position, are not necessarily reflected by a game's final score. Therefore, each match is observed by a total of six judges that award positive scores for actions that express team play and negative scores for actions that negatively affect the team performance.

One major manifestation of cooperation in soccer is passing. Hence, a robot that plays a pass as well as a robot that receives a pass should be awarded with positive scores. There are no fixed scores for these actions; instead, judges are allowed to assign scores within a range as this allows the judges to differentiate between almost accidental passes and intended passes that provide a huge benefit to the robot's team. Although the action is initiated by the passing robot, the receiver also earns a positive score as it enabled this situation by intelligent positioning.

In normal SPL games, robots are penalized for pushing their opponents but pushing their teammates is ignored and considered as the team's own fault. However, in the Drop-in Competition, pushing a teammate results in a negative score as it is considered to be a disadvantage for the teammates.

For singular actions that are not passes or pushes of teammates but nonetheless influential for the whole team, the judges are allowed to give an unclassified bonus or penalty. This could be positive scores for helpful support positioning or negative scores for blocking teammates or even stealing the ball from them. In any case, the judges are required to write down a justification of the unclassified bonuses given by them to gain insights for possible future rule adaptations.

For each half, each robot receives a positive or negative score for its game participation, i. e. its overall contribution to the team's performance. The score for game participation was not part of the original rule set but was introduced after the first drop-in games at local RoboCup competitions. During these games, it turned out that almost passive robots that stood and observed the game most of time received better average scores than robots that had been highly active and that tried to contribute to the game. The latter often accidentally pushed their teammates or sometimes stole a ball from a teammate and thus received quite negative scores. To avoid teams converging to a *Do not cause harm* tactic, the score for game participation was added to directly encourage robots to attempt to contribute to the game.

Observed Behavior	**Score Range**
Pass to a teammate	+1 – +4
Receiving a pass	+1 – +4
Pushing a teammate	-2
Unclassified bonus or penalty	-2 – +2
Game participation (once per half)	-10 – +10

Table 1: Possible scores (and their ranges) that can be awarded by judges. The *Unclassified bonus or penalty* is capped at -10/+10 per half.

In each half, the six judges are divided into two groups of three with one group observing each team. At halftime, the judges switch teams. This procedure results in six scores from six different judges per robot per game. Each judge has to observe a maximum of five robots at any point of time and fill out a score sheet[1]. The possible score ranges of these sheets are listed in Table 1. The judges were encouraged to read about how the scoring criteria should be applied in the SPL rulebook (RoboCup Technical Committee 2014b) before serving as judges.

3.4 Organization of the Competition

Drop-in Competition games are played 5 vs. 5 just like normal SPL games. In each game, all 10 robots on the field always originate from different teams to avoid any biases in the overall scores. During a tournament, multiple games are played and the scores of each robot become averaged over all its games. To achieve meaningful scores that reliably reflect the drop-in capabilities of a single robot, it is best to play as many games as possible. It is also best to play with as many different teammates and opponents as possible.

For the RoboCup 2014 Drop-in Competition, we were able to schedule 15 drop-in games into the schedule. As initially 25 teams registered for the Drop-in Competition, every robot participated in 6 different matches. The pairings were randomly generated by a computer program that was also used for 3D Soccer Simulation League drop-in games (Algorithm 1 in (MacAlpine et al. 2014)). To consider some organizational details, minor adaptations of the program as well as some manual adjustments of the schedule had been applied. Not all teams who registered for the Drop-in Competition actually showed up — two teams missed all of their games and some teams missed a few of their games — but their spots remained empty, which resulted in not all games being played 5 vs. 5. This unfortunately did affect games,

[1] https://www.informatik.uni-bremen.de/spl/pub/Website/Downloads/dropInScoreSheet2014.pdf

but not necessarily significantly because their absences were spread across teammates and opponents.

This kind of schedule led to the following characteristics:

- A team does not play with every other team. The computer program's solution for such a prerequisite was to hold 27 games in total. Within the context of a RoboCup competition, which runs a huge number of normal games as well as multiple technical challenges, scheduling such a number is not realistic.
- During the tournament, each team had at least 18 and up to 22 different teammates.
- Players will not play with each other more than 3 times.
- During the tournament, each team has at least 17 and up to 20 different opponents.

In addition to generating games with robots from 10 different teams, each game requires four referees (preferably from two different teams) as well as six judges (preferably from six different teams). Judges and referees are always selected from teams that are not playing in a match. Overall, running a single Drop-in Competition game involves people from up to 18 different RoboCup teams.

To avoid any pre-coordination, the exact assignment of robots to teams and numbers to robots was announced as late as possible. Initially, only the time slots of the matches were announced to allow teams to prepare for these time slots. The exact time of the announcement of each games' details varied between 30 minutes and multiple hours, depending on each day's overall schedule and the need to inform everybody in time to avoid any misunderstandings. In addition to the late announcement, all teams were told to refrain from planning any pre-coordination.

4 Drop-in Player Strategies

All teams participating in the 2014 Drop-In Player Competition were asked to submit a short description of the strategy they used in the competition. In total, 17 out of 23 participating teams submitted a description. The original texts were made public at (RoboCup Technical Committee 2014a) to allow the teams to learn from each other and to provide a better overview of the current status of the competition, i. e. to give a set of answers to the question: *What is the difference between playing with and without pre-coordination?*

4.1 Communication and Coordination

As described in Section 3.2, all robots within a team are connected by a wireless network and are able to send standardized messages to each other. In theory, these messages should be a valuable source of information for a robot to coordinate with its teammates. In practice, it is not guaranteed that a proper communication can be established because

- not all robots actually send messages
- not all robots fill all standard message elements
- some robots compute wrong data, likely as a result of delocalization or false positive ball observations.

In their strategy description, more than half of the teams do not mention these problems or explicitly state that they trust their teammates. However, seven teams mentioned that they do not accept all communicated messages:

- Three teams state that they discard *most* of the information that they receive. However, they do not explain any criteria for accepting or dropping messages (or parts thereof).
- One team did not implement the communication protocol at all.
- One team sends messages but discards all messages that it receives.
- Two teams implemented approaches to determine the reliability of their teammates by checking the plausibility of the transmitted information and the teammates' ability to fulfill their announced roles, respectively.

As described in the next section, this limited communication affected the chosen strategies in multiple cases.

4.2 Typical Player Behaviors

Regarding the roles chosen by the playing robots, there appears to be one strategy applied by the majority of the teams:

Try to play the ball, if it is close and/or no other robot wants to play to the ball. Take a supporting position otherwise.

In many cases, the decision to go to the ball depends on the communicated positions and intentions of the teammates. The chosen supporting positions vary from simple strategies like *Staying close to the ball* to more complex computations involving the positions of all other teammates. These strategies are, as mentioned by multiple teams, often the same ones as used for their normal games.

However, some of the teams that accepted only few or even no messages from their teammates switched to a different strategy to avoid conflicts, and thus possible negative scores, with teammates that also want to play the ball. They position their robots at more or less fixed positions on the field, e. g. a defensive position inside the own half or somewhere close to the field's center, and let them wait for the ball to appear in their proximity. If this happens, the robots start to kick or to dribble towards the opponent goal. Otherwise, the just remain at their position and track the ball.

One role that was only mentioned in few descriptions and rarely seen in actual games was the one of the goalkeeper. There were teams that actively avoided this role as they had the impression that the current scoring scheme, which awards activity, disadvantages goalkeepers. Well-programmed goalkeepers would likely receive positive scores — but teams with well-programmed goal keepers likely felt they would assist the team more in other roles.

5 Results and Analysis

In the SPL Drop-in Player Competition, twenty-three teams participated in fifteen full-length games. As discussed in Section 3.3, the overall winner of this competition was determined via two metrics: average goal difference and average human-judged score. The two scoring metrics were normalized and added up as specified in the official SPL rulebook(RoboCup Technical Committee 2014b) to determine the overall winner. The results are displayed in Table 2 and analyzed in the remainder of this section.

One of the goals of the SPL Drop-in Player Competition is for a team comprised of the top five drop-in players to be able to play comparably to the winner of the main soccer SPL team competition. At RoboCup 2014 we held the first of these ‘all-star’ games where robots from B-Human, HTWK, Nao Devils, TJArk and Berlin United played together as an ad hoc team against the 2014 SPL champion rUNSWift in a full-length normal SPL game. The result was 4-2 in favor of rUNSWift, but the relative closeness of the result shows that the ad hoc team did well in this first ‘all-star’ game.

Agents designed for the Drop-in Player Competition should be adept at reasoning about their teammates’ abilities and intentions and responding in such a way that helps their team the most. The authors carefully designed the scoring metrics presented in Section 3.3 to reward agents for being good teammates and not for having better lower-level skills. Despite having a standard platform in the SPL, some teams have designed substantially superior walk engines and kick engines that could make them seem to be a better teammate solely because they are more skillful. Hence, this is one of a reasons why the SPL Drop-in Player Competition uses both human judge scores and score differential in order to select an overall winner.

In the 2013 technical challenges, both of the simulation leagues used only goal differential to determine the winner of their challenges. However, simulation leagues can run many full-length games in parallel across multiple machines, whereas the SPL is limited by field space, team member time, referee time, and availability of team robots. Hence, the SPL can not run nearly as many games. As such, we also use judge scoring to help offset the scoring noise potentially caused by a limited number of games.

So far in this section we have considered the reasons to use human judges in the SPL Drop-in Player Competition. However, we do believe that goal differential is also very important because it embodies the main aspect of being a good teammate — helping your team win.

With this in mind, let us further analyze the judge scores, goal differential scores, and overall ranking when compared to rankings in the main SPL soccer competition.

5.1 Analysis of Judge Scores

When looking at the judge scores in Table 2, the fact that UTH-CAR had a substantially worse judge score than any other team stands out. This substantially lower judge score had a large impact on the results of the competition because it caused 22/23 teams to have a normalized judge score (Judge Norm in Table 2) of greater than 54 and 17/23 teams to have a normalized judge score of greater than 70. This caused the judge scores to have a weaker influence on the overall Drop-in Player Competition rankings than anticipated.

UTH-CAR had such a low judge score because their robot was often inactive or not on the field, and hence was rated as a poor teammate. Other teams, such as UChile and SPQR, also failed to put an active robot on the field for some of their games, and hence received low judge scores.

Scoring by human judges is inherently subjective and inconsistent. Human judges were selected from teams not participating in the current match, but often the human judges were distracted or confused about when to award bonuses and penalties to individual robots. Hence, despite averaging across six judges for each game, the judge score were likely not a very accurate representation of each robot’s ability as a teammate. We believe human judging can be improved upon for future competitions by a combination of better scoring criteria and improved judge training.

5.2 Analysis of Goal Differential Scores

Average goal differential across a robot’s games is a non-subjective measure of how well the robot’s team did. If enough games are run — as is possible in simulation leagues — then average goal differential would be an excellent stand-alone metric for how good of a teammate a robot is. However, with limited games and inconsistent opponents, average goal differential is a noisy estimate. Additionally, average goal differential is unable to differentiate whether a robot is scoring well because of good teamwork or because of superior skills.

One interesting aspect of the goal differential scores is that some teams, such as UTH-CAR, failed to consistently put a robot on the pitch (and received the worst judge score because of this) yet managed not to obtain one of the worst goal differential scores. Indeed, UTH-CAR’s goal differential ranking was tied for 16th. Hence, despite not even being on the pitch in many games, UTH-CAR did decently in terms of goal differential.

5.3 Comparison of Competition Rankings

It can be assumed that teams that perform well in the main SPL RoboCup competition generally have better low-level skills than those who perform poorly. Hence, we can compare each team’s Drop-in Player Competition rank (Drop-in Comp Rank in Table 2) to their main SPL competition rank (Main Comp Rank in Table 2). Specifically, Main vs Drop-in Rank in Table 2 shows how much better or worse each team placed in the Drop-in Player Competition as compared to in a the main competition.

In general, better teams in the main competition did tend to perform better in the Drop-in Player Competition — only one team that finished tied for 13th in the main competition was in the top 9 teams in the Drop-in Player Competition. Interestingly, some teams who performed very well in the main competition, namely MRL and UChile, finished in the bottom three teams for the Drop-in Player Competition. This result suggests that solid low-level skills and deployment of normal game code will not necessarily yield success in the Drop-in Player Competition, but that teams with good low-level skills and solid Drop-in Player Competition teamwork protocols will likely perform well.

6 Related Work

Although multiagent teamwork is a well-studied area, most research addresses the problem of coordinating and com-

Team	Country	Judge Avg	Judge Norm	Judge Rank	Goal Diff Avg	Goal Diff Norm	Goal Diff Rank	Drop-in Comp Score	Drop-in Comp Rank	Main Comp Rank	Main vs Drop-in Rank
B-Human	Germany	4.72	100	1	1.33	100	1	200	1	3	+2
HTWK	Germany	1.28	83.04	6	1.00	89.47	2	172.51	2	2	0
Nao Devils	Germany	1.61	84.68	4	0.67	78.95	T4	163.63	3	T5	+2
TJArk	China	2.17	87.41	3	0.50	73.68	T6	161.10	4	T9	+5
Berlin United	Germany	-0.58	73.87	12	0.67	78.95	T4	152.82	5	T5	0
DAInamite	Germany	0.08	77.15	9	0.50	73.68	T6	150.84	6	T13	+7
UPennalizers	USA	0.67	80.03	8	0.33	68.42	T8	148.45	7	T9	+2
Austrian Kangaroos	Austria	-2.90	62.45	19	0.83	84.21	3	146.66	8	T9	+1
rUNSWift	Australia	3.00	91.52	2	-0.17	52.63	T13	144.15	9	1	-8
Cerberus	Turkey	0.72	80.30	7	0.00	57.89	T10	138.20	10	T13	+3
Northern Bites	USA	-1.81	67.85	17	0.33	68.42	T8	136.27	11	T13	+2
NTU RoboPAL	Taiwan	1.61	84.68	4	-0.50	42.11	T16	126.78	12	T5	-7
UT Austin Villa	USA	-1.28	70.45	16	-0.17	52.63	T13	123.08	13	T13	0
HULKs	Germany	-1.83	67.72	18	-0.17	52.63	T13	120.35	14	T13	-1
UnBeatables	Brazil	-3.36	60.19	20	0.00	57.89	T10	118.09	15	-	-
RoboCanes	USA	-1.06	71.55	14	-0.50	42.11	T16	113.65	16	T13	-3
Philosopher	Estonia	-0.25	75.51	11	-0.67	36.84	19	112.36	17	T13	-4
Edinferno	UK	-0.08	76.33	10	-0.83	31.58	20	107.91	18	T13	-5
MiPal	Australia/ Spain	-0.94	72.09	13	-1.00	26.32	21	98.41	19	-	-
SPQR	Italy	-8.00	37.35	22	0.00	57.89	T10	95.24	20	T9	-11
MRL	Iran	-1.22	70.73	15	-1.33	15.79	22	86.52	21	T5	-16
UChile	Chile	-4.50	54.58	21	-1.83	0.00	23	54.58	22	4	-18
UTH-CAR	Mexico	-15.58	0.00	23	-0.50	42.11	T16	42.11	23	-	-

Table 2: Scores for the 2014 SPL Drop-in Player Competition (listed from best to worst).

municating among teams that are created to work together and hence share a common coordination framework (Tambe 1997; Grosz and Kraus 1996; Horling et al. 1999). Ad hoc teamwork, on the other hand, addresses multiagent teamwork in which all of the coordinating agents do not share a common coordination framework. Liemhetcharat and Veloso focused on how to select agents to form ad hoc teams based on each agent's individual characteristics and interactions with its teammates (Liemhetcharat and Veloso 2011). Barrett *et al.* present empirical evaluations of various types of ad hoc agents when joining coordinated teams of unknown agents in the Pursuit domain (Barrett, Stone, and Kraus 2011). Jones *et al.* present a treasure hunt domain for evaluating ad hoc team performance and present a simple implementation of a team that can search for treasure in such a domain (Jones et al. 2006).

In the robot soccer domain, Bowling and McCracken (Bowling and McCracken 2005) propose methods for coordinating an agent that joins an unknown, pre-existing team. In their work, each ad hoc agent is given a playbook that differs from the playbook of its teammates. The teammates assign the ad hoc agent a role, and then react to it as they would any other teammate. The ad hoc agent analyzes which plays work best over hundreds of simulated games, predicts the roles its teammates will adopt in new plays, and assigns itself a complementary role in these new plays.

Although related, none of this previous research besides the 2013 drop-in challenge (MacAlpine et al. 2014) has been evaluated on real robots programmed by various organizations from around the world in a truly ad hoc teamwork setting. This paper expands on the 2013 drop-in challenge by substantially increasing both the number of teams participating and the number of drop-in games held.

7 Conclusion

The Drop-in Competition in the SPL matured at RoboCup 2014 into a useful testbed for cooperation without pre-coordination. With the SPL being a standard platform league, and with options existing for teams to just compete in the SPL Drop-in Competition at RoboCup, this testbed is open and approachable for multiagent systems researchers looking to work on ad hoc teamwork in a robotics domain. The SPL plans to continue this competition for the foreseeable future. The league's Drop-in Player Competition goal is to be able to create a team comprised of the top five drop-in players that can play comparably to the champion team.

As with any competition, there are always improvements that can be made. One aspect of the competition that can be improved is the human judging criteria. In particular, how can criteria be created that allows many different judges to consistently and accurately judge a robot's teamwork? Additionally, how can the competition rules be written such that outliers do not drastically affect the normalization process?

8 Acknowledgements

Katie Genter and Peter Stone are part of the Learning Agents Research Group (LARG) at UT Austin. LARG research is supported in part by NSF (CNS-1330072, CNS-1305287), ONR (21C184-01), and AFOSR (FA8750-14-1-0070, FA9550-14-1-0087).

References

Barrett, S.; Stone, P.; and Kraus, S. 2011. Empirical evaluation of ad hoc teamwork in the pursuit domain. In *AAMAS'11*.

Bowling, M., and McCracken, P. 2005. Coordination and adaptation in impromptu teams. In *AAAI'05*.

Grosz, B. J., and Kraus, S. 1996. Collaborative plans for complex group action. *Artificial Intelligence* 86(2):269–357.

Horling, B.; Lesser, V.; Vincent, R.; Wagner, T.; Raja, A.; Zhang, S.; Decker, K.; and Garvey, A. 1999. The TAEMS white paper.

Jones, E.; Browning, B.; Dias, M. B.; Argall, B.; Veloso, M. M.; and Stentz, A. T. 2006. Dynamically formed heterogeneous robot teams performing tightly-coordinated tasks. In *ICRA'06*, 570–575.

Liemhetcharat, S., and Veloso, M. 2011. Modeling mutual capabilities in heterogeneous teams for role assignment. In *IROS'11*.

MacAlpine, P.; Genter, K.; Barrett, S.; ; and Stone, P. 2014. The RoboCup 2013 drop-in player challenges: Experiments in ad hoc teamwork. In *IROS'14*.

RoboCup Technical Committee. 2014a. 2014 drop-in player strategies. Only available online: `http://www.informatik.uni-bremen.de/spl/bin/view/Website/DropinStrat2014`.

RoboCup Technical Committee. 2014b. RoboCup standard platform league (nao) rule book. Only available online: `http://www.informatik.uni-bremen.de/spl/pub/Website/Downloads/Rules2014.pdf`.

Stone, P.; Kaminka, G. A.; Kraus, S.; and Rosenschein, J. S. 2010. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *AAAI'10*.

Tambe, M. 1997. Towards flexible teamwork. *Artificial Intelligence Research* 7(1):83–124.