Team sports for Game AI benchmarking revisited

MAXIM MOZGOVOY, The University of Aizu, Japan

MIKE PREUSS, LIACS, Universiteit Leiden, Netherlands

RAFAEL BIDARRA, Computer Graphics and Visualization Group, Delft University of Technology, Netherlands

Sport games are among the oldest and best established genres of computer games. Sport-inspired environments, such as RoboCup, have been used for AI benchmarking for years. We argue that, in spite of the rise of increasingly more sophisticated game genres, team sport games will remain an important testbed for AI benchmarking due to two primary factors. First, there are several genre-specific challenges for AI systems that are neither present nor emphasized in other types of games, such as team AI and frequent re-planning. Second, there are unmistakable non skill-related goals of AI systems, contributing to player enjoyment, that are most easily observed and addressed within a context of a team sport, such as showing creative and emotional traits. We analyze these factors in detail and outline promising directions for future research for game AI benchmarking, within a team sport context.

Additional Key Words and Phrases: Sport games, Team sports, AI benchmarking, Game AI

1 INTRODUCTION

Historically, Artificial Intelligence (AI) researchers have often relied on popular games as testbeds for evaluating new algorithms and approaches. Discussing a famous adage *"chess is a Drosophila of AI"*, J. McCarthy wrote: *"To computer scientists in general I have only one wish to express: let there be more of these Drosophila-like experiments; let us create some more specific examples!"* [15].

Several factors may contribute to the success of a particular testbed. For example, in the case of chess one may note that the game is easy to setup, it has a wide appeal both to general public and to researchers themselves, it poses problems that are perceived as generalizable beyond the game world, and even allows independent study of isolated aspects of the game such as endings.

We can say that most *team sports* have similar features: they are popular among the general public, they are easy to setup, and they require the participating team members to exhibit both athletic abilities and a certain level of tactical and strategic thinking. Although various definitions of team sport can be given, we propose to abide here by that put forward in a recent survey: "a game that typically involves two teams playing against each other, each composed of a set of players with their individual roles and abilities" [2]. If further specification is desired on what makes up a team in contrast to a group, we may go with the definition given by Thompson (from a management perspective) [32]: "a team is a group of people who are interdependent with respect to information, resources, and skills and who seek to combine their efforts to achieve a common goal". What is not explicitly said here but clear from the context is that there must be ways of communication between the team members. If we add communication to the team sport definition above, we may get a more complete definition: *team sports are games that involve at least two teams playing against each other, each composed of a set of players with their individual roles and abilities who cooperate by means of communication in order to win the game.*

These features make team sports an interesting challenge for a game AI system and we, therefore, propose to consider them as a promising testbed for AI benchmarking. In the remainder of this paper we outline our vision on the topics involved.

Before suggesting to focus AI benchmarking more on the team sports area, let us first have a look at how the competition landscape currently evolves. Without doubt, game-related benchmark environments have recently propelled AI progress and led to a huge impact in the media, thereby contributing to the current AI hype.

The game of Go was already an unofficial AI benchmark when chess lost much of its appeal for AI researchers as a result of DeepBlue's victory over Kasparov in 1997. Whereas Silver *et al.* [28] already provided an approach that was able to beat professional human Go players, the old Atari console environment was used to invent *Deep Reinforcement Learning* (DRL) [17, 18]. It soon turned out that most of the games of the the so-called *Atari learning environment* (ALE) are relatively easy to deal with by this new technique, even playing many games with super-human performance was quickly achieved. In search for new challenges, research turned to *real-time strategy* (RTS) games.

StarCraft was established as a benchmark and competition environment in 2010 [20], and was still considered very challenging when DeepMind and Blizzard teamed up to provide the SC2LE environment for StarCraft 2 in 2017 [36]. Following the reports of the AlphaStar team of DeepMind [35], it seems that RTS games are not yet completely done, but getting in reach. This is suprising on one hand, because most researchers had expected it to take years to get that far. On the other hand, there is higher emphasis on generalizability now: we are not satisfied with an AI system that can be used for a single game and in previously seen environments anymore; we expect the AI shall be able to transfer knowledge between situations, environments, and eventually games.

An interesting development in this direction is the Obstacle Tower Challenge [11] that features almost endless generated mazes that gradually get more difficult. Generalization in game mechanisms and visuals is key here, but there is only a single avatar that is completely autonomous, thus interaction happens only with more or less static objects.

Another aspect, the team play between different AI systems, is emphasized in the cooperating OpenAI Five[1] *Multiplayer online battle arena* (MOBA) game bots. The number of bots in a team is limited to five, but they clearly need to work together well in order to win a game — a capability that is also fundamental for team sports AI. We, therefore, also compare the properties of MOBA, RTS and team sports games and the resulting challenges for AI later on.

With this work, we highlight the need to define good benchmark environments that integrate *team AI*, here understood as multiple game AIs that interact among themselves and with humans in order to reach a common goal. For this, team sport games provide an excellent context with numerous advantages.

We start this endeavour with reviewing the differences between real-world and virtual team sports (Section 2), before discussing the case of RoboCup (Section 3), a famous team sports AI environment that was actually not conceived as a game. As said before, user enjoyment is an important aspect also for AI competitions, and it is addressed in Section 4. After discussing the interplay of strong and fun game AI in Section 5, we also look at the tactics and strategy perspective (Section 6), before more generally collecting the challenges of sports AI, also in comparison with MOBA and RTS (Section 7) and ending with conclusions.

2 REAL-WORLD AND VIRTUAL TEAM SPORTS

There is no universal definition of "team sports". According to Collins Dictionary, a team sport is *"a sport in which teams play against each other"* [4]. The current Wikipedia article expands this definition as follows: *"A team sport includes any sport where individuals are organized into opposing*

---

[1]https://openai.com/blog/openai-five-benchmark-results/

teams which compete to win. Team members act together towards a shared objective. This can be done in a number of ways such as outscoring the opposing team" [37].

For some sports, one may wonder to which extent they should be considered full-fledged team sports. For example, in rowing, relay racing or swimming, there *are* competing teams of athletes, but little communication is required among team members and virtually no contact with the opposing team.

More interesting cases are sports in which teams take on distinct roles, and follow different objectives in different phases of the game, like fielding team and batting team in baseball, or when each team consists of several groups with immutable roles as in quidditch[2], where two largely independent but somehow interacting games (chaser and beater) take place on the same field, extended by a third one (seeker) during the end phase of the game.

Having noted these scenarios, let us narrow down the subsequent discussion to the types of sports in which both teams have the same structure and follow the same goal as, for example, in ice hockey, basketball, soccer, volleyball and many other similar games.

Sport games are also a genre staple of the video games industry. Since *Tennis for Two* (1958), virtual renditions of sport games spread literally to every gaming platform available. Obviously, it is not possible to capture *all* aspects of a physical athletic activity on a computer, and different game projects emphasize different elements of real-world sport events. As extreme examples, departing from a somewhat standard formula of a sport video game, we can mention *Subbuteo* (1990), representing soccer as a turn-based billiard-like strategy, *Captain Tsubasa* (1988), emphasizing role-playing elements of soccer and player relationships, and a number of games like *Super Mario Strikers* (2005) or *Nintendo World Cup* (1990), with unrealistic "fun" elements such as powerful supershots, players' special abilities, and extreme weather conditions. Even within the same franchise, developers sometimes create games that emphasize "simulation" or "fun" aspects, such as in case of EA's *NBA Live* vs. *NBA Jam* or *NHL* vs. *NHL Slapshot*. Therefore, the discussion around "team sports as a game AI testbed" requires first deciding which aspects of team sports should be investigated, and what are the ultimate goals.

3 THE CASE OF ROBOCUP

Fortunately, we have a great reference point: the event series organized by the RoboCup Federation [23]. One may wonder why we are bringing the topic of team sports AI benchmarks in spite of the existence of an established RoboCup framework. The main reason for it is the very specific agenda of the RoboCup Federation, which ultimate goal is stated as follows: *"by the middle of the 21st century, a team of fully autonomous humanoid robot soccer players shall win a soccer game, complying with the official rules of FIFA, against the winner of the most recent World Cup"* [24].

On the way to the accomplishment of this grand vision, a number of smaller goals have to be achieved. Thus, RoboCup competitions consist of several independent events, each focusing on a few relatively isolated subproblems, such as hardware challenges, computer vision, and team-based behavior strategies. In practice it means separation into "hardware" and "software" leagues, and organization of different tournaments for different types of robots (including separate competitions for 2D and 3D virtual software-simulated robots).

These tournaments are deliberately designed to support the grand vision of RoboCup organizers, and to ensure smooth knowledge transfer between independent RoboCup events. For example, software robots must deal with the same type of constraints as their hardware counterparts, such as limited visibility, noisy sensory input, imprecise locomotion, and low-level actions (for instance, an agent may not dribble the ball: it has to program each ball kick independently).

---

[2]https://en.wikipedia.org/wiki/Quidditch_(sport)

Such setup is perfectly reasonable given the ultimate goal of RoboCup competitions. However, RoboCup environments, such as 2D Simulation League [25], were not designed to resemble computer games. While one can connect mouse and keyboard to RoboCup software and take control of one of the 22 players, this experience will be significantly different from a typical sport video game session.

Let us focus on the 2D Simulation League, representing the most video game-like environment among RoboCup events. It is striking that the pool of successful 2D Simulation League participants is very stable. During the last 5 years, only 7 different teams managed to reach a top 3 position[3] (the total number of participants in each of these events ranged from 13 to 19 teams). All these teams were established at least 6 years prior to their prize-winning seasons.

In addition, we also notice a general decline in participation: 24 teams played in each of five consecutive seasons from 2000 to 2004, and since that period no competition could attract more than 20 teams.

We believe these observations can be at least partially explained by the sheer amount of work required to establish a decent team: high entrance barriers work against aspiring contestants. In general, a skillful RoboCup 2D Simulation League AI system must be able to:

- Efficiently and reliably translate high-level decisions (such as "run with a ball for 10 meters") into a series of low-level kick and dash actions.
- Reason on the basis of noisy and limited audiovisual sensory data.
- Handle teammate-teammate and teammate-coach messages using limited communication channels.
- Efficiently distribute roles between the teammates according to their skill profiles generated by the RoboCup server at the beginning of a match.
- Adjust its strategy if some player is removed from the field by a referee.
- Be good at playing set pieces (kick-ins, free-, penalty- and corner-kicks).

While these diverse abilities are relevant to the field of game AI, it is difficult to develop all of them within a small group having expertise in just one specific subfield of AI research. A public release of the award-winning Helios team's *agend2d* codebase [1] motivated around 80% of other teams to abandon their own developments and switch to *agent2d* [21]. This process is seen by different authors as either "greatly beneficial" [22] or "detrimental" and leading to lower diversity of teams [14]. Furthermore, RoboCup was never designed as a *game* world; instead, it was conceived as downgraded (software-only) robot training environment. This opens up an additional perspective for our discussion.

4 THE ROLE OF FUN IN GAME AI COMPETITIONS

Which game environments are good for AI competitions? Discussing a related topic of organizing such contests, Togelius [33] gives the following advices *"choosing a fun game"* (*"it also helps if the game is famous"*, he adds) and *"making it really easy to get started"*. Applying this to our domain, we might add as well that it definitely helps if the respective real-world sport is popular worldwide.

RoboCup apparently relies on another formula. Analyzing a 20-year long story of RoboCup, Ferrein and Steinbauer [7] note *"the atmosphere of some three thousand robot enthusiasts"* and *"fascinating outreach to the general public"*, and emphasize the interdisciplinary nature of participants' work, community building and team building efforts, and a chance for the participants to solve real robotics problems. In other words, it seems that RoboCup is backed first and foremost with

---

[3]WrightEagle, Helios, Gliders (Fractals), Cyrus, FRA-United, Oxsy, and MT. See https://ssim.robocup.org/soccer-simulation-2d/2d-awards/

people's interest to physical robots, while "game-like" aspects of this event (especially in case of simulation leagues) play a somewhat secondary role.

The notion of what constitutes fun is largely subjective, but we can at least remark that, while it is *possible* to play RoboCup 2D Simulation League matches between human-controlled teams, people do not really do it; and, in any case, the gameplay of such a match would radically differ from any successful commercial soccer game. It is, therefore, highly likely that (i) RoboCup is not a very fun *video game*, and (ii) there is a niche for sport games in *game* AI competitions, which is not filled by RoboCup by design. One of the candidates for filling this niche could be the game Rocket League [34] which combines ball game and car racing fun elements and is already played by different bots in teams of 3.

The role of fun in a game is clearly broader than just to be an instrument for keeping the public interested in a certain AI competition. But it is one of the factors, shaping the game world, directly affecting the design goals of a game AI system, and the principles of its subsequent evaluation. Long-standing popularity of a certain game genre proves that the core game mechanics is both fun (*game-wise*) and challenging enough to continue fueling interest of both spectators and participants.

For team sports, the battle of team strategies has always been a major part of spectator enjoyment. Sport fans appreciate teams showing both spectacular and efficient play, and the balance between these goals is not easy to achieve. In addition, professional regulating bodies monitor actual developments of common game tactical patterns, and adjust the rules to keep competitions appealing to the audience and fight against degenerate strategies [12]. Given the long history of most popular team sports, one can assume that current rules are well balanced, and encourage inventive, non-trivial team play. For example, the early offside rule of soccer, introduced to prevent "goal hanging" (a degenerate yet efficient tactical pattern) [3], underwent significant revisions in 1925 and 1990, and was last refined in 2005 — every time to encourage attacking play and to limit the abuse of defensive "offside traps" [38, 39]. Similarly, a back-pass rule, preventing the goalkeeper from handling the ball received from a teammate, was introduced in 1992 and extended in 1997 as an anti-time-wasting measure [8].

These considerations make us believe that team sport games are an appealing choice as a game AI testbed: they are fun, spectacular, well-balanced, reasonably complex, easy to setup, rely on simple rules, and enjoy a wide fan base.

5   STRONG GAME AI AND FUN GAME AI

While games can be used and are actually used to benchmark general AI technologies, there are important features, characterizing *game* AI systems. It is generally presumed that the main point of a game is to win, and thus the best AI system in a typical AI competition is the system that wins.

However, the purpose of game AI is not necessarily to be *strong*. According to Dill [6], *"The one thing that is universally true is that games are about creating a particular experience for the player — whatever that experience may be. The purpose of Game AI (. . .) is to support that experience."* Thus, depending on a particular game, the goal of a *good* AI system might well involve being strong, predictable, erratic, friendly, hostile, and so on [30]. In other words, a good game AI testbed should support, at least, theoretically, various possible goals for AI-controlled characters.

The goal of strong AI development for games like chess or Go can be considered achieved, since computers are able to defeat even the best human players. We can expect that more game genres will be added to this list in the nearest future. For example, a recent work by Oh *et al.* [19] discusses the development of an AI system, able to defeat professional human players in a modern fighting game.

It is difficult to say how good modern AI methods are in playing sport games. Michael and Obst [16] observe that AI teams of RoboCup 2D Simulation League play better than human teams;

they remark, however, that RoboCup is not designed to be played by people, which supposedly affects their performance. In any case, there are independent goals of designing a strong AI system, and an AI system that contributes to the overall user enjoyment. Arguably, the latter task is even more important for the needs of practical game development.

Team sports are a good testbed for investigating such non-skill related traits of AI systems as well. There is extensive literature on factors making team sports exciting for both athletes and spectators, for which there are many good examples [5, 26]. Likewise, there is a general understanding of what constitutes fun in the context of an AI system for a sports video game. In particular, we often observe that people prefer playing against other people, because people behave in a certain "human-like" way that is perceived as inherently enjoyable [29]. In conclusion, striving for a human-like behavior can be a legitimate goal for a sport game AI system, as important and challenging as a highly skilled behavior.

6 TACTICS AND STRATEGY IN SPORTS

An obvious problem with the choice of team sports as an AI benchmark lies in the fact that they require certain physical abilities and athletic skills in addition to tactical and strategic decision making. Since these physical aspects cannot be captured within the game framework, it is necessary to investigate the extent to which sport team behaviors pose a significant challenge for AI. This, in turn, can be further split into questions as: 1) what is the role of tactics and strategy in real-world team sports?; 2) what is the complexity of actual decision making in real-world team sports?; 3) how well can these experiences translate into a video game setting (see Fig. 1)?
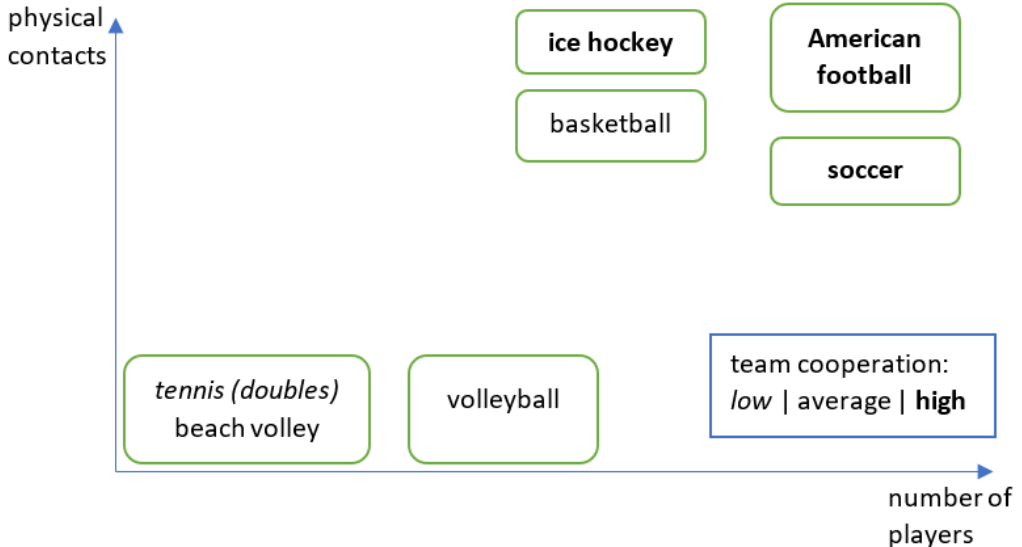

Fig. 1. Features of particular team sports in 3 dimensions: number of players, physical contacts, importance of team cooperation. Games written in boldface require the highest team cooperation levels. Here, American football has the highest levels in all three dimensions.

Obviously, the contribution of tactical/strategic factors highly depends on a particular sport. However, there is a plethora of works dedicated specifically to these game elements in a large

Table 1. **Control** property. Are there multiple autonomous agents?

Team sports AI	Mostly yes; for RoboCup, 12 per team (11 players + the coach agent directing the overall strategy).
MOBA AI	Yes, but mostly at the end of the game — the roles are quite fixed at the beginning, and there is not much coordination.
RTS AI	Central control is possible (but not enforced) for lower-level decisions; strategic-level decisions are centrally controlled.

Table 2. **Unit diversity** property. Does the AI have to find suitable ways to combine different units or to counter the enemy strategy?

Team sports AI	Average; units can be quite different, but, in principle, most units are able to replace other units.
MOBA AI	High; units are definitively different and need to be combined in suitable ways; only a small subset of possible unit types is used in each game.
RTS AI	High, but there is less diversity in abilities (special powers) than in MOBA.

variety of team sports [9, 10, 31, 40]. In turn, good game strategies are not easy to construct. For instance, soccer history is characterized by a series of paradigm shifts in building up team strategies, as Wilson remarks in his book solely dedicated to the history of soccer tactics [38]: *"football is not about players, or at least not just about players; it is about shape and about space, about the intelligent deployment of players, and their movement within that deployment."*

The differences in player abilities impose additional constraints on coach decisions: one has to devise an efficient tactical scheme, applicable in the given context, in a match between two specific teams consisting of specific players. We can observe this phenomenon even in relatively uncomplicated games, such as beach volleyball, where each team consists of two players, and there is no physical contact with the opponents. Koch and Tilp [13] show how differences in physical characteristics between male and female athletes (among other factors) encourage them to prefer distinct techniques in order to win.

Speaking of computer game renditions of sports, it seems that the goal of the mainstream AAA projects, such as EA's FIFA, is *"to make the games even closer to the actual game, that is, to make the computer game converge with the sport"*, as observed by Sicart [27]. Interestingly, in his analysis of the differences between the real soccer and FIFA'12, the most salient feature was FIFA's AI system, controlling the players and the referee. Sicart is generally satisfied with simulation of physical aspects of the game, but believes that AI falls short in its understanding of soccer. While real soccer rules leave enough room for referee's interpretation, AI referees of FIFA'12 are scripted in a deterministic way, leaving no space for ambiguity. Similarly, AI-controlled players behave in a logical yet predictable manner, and even superstar players honored for their creativity and tactical vision (like Messi) do not exhibit such abilities in the computer game.

These limitations actually have deep implications for the players. The rigid, predictable nature of AI decision-making encourages people to exploit this knowledge and build their strategy around it. As Sicart puts it, *"FIFA players (…) need to learn how to think procedurally, how to decode the technical implementation of a known set of rules, tactics, and player characteristics, and apply this way of thinking to ways of playing the game."* This observation reinforces our proposition that sport game AI is still far from satisfactory, and requires further research.

Table 3. **Agent roles** property. Does the game impose specific roles the agents have to fulfill?

Team sports AI	Yes, but there is a variety of "play systems" with slightly different roles (in general, there are offensive and defensive roles, and often dedicated goalkeepers).
MOBA AI	Quite diverse roles and very diverse agent templates (heroes); the choice of heroes alone can have a significant impact on team performance.
RTS AI	Roles are not obvious in RTS, except in cases of clear specialization (workers / army units / air units); units have little individuality and mostly count as a part of their squad.

Table 4. **Action spaces** property. Does the agent have many micro-level actions available?

Team sports AI	Yes, several: movement, different ball actions (shooting, dribbling, header — most of them are parameterized); also specific actions for interaction with other agents (e.g., fouls and tackles).
MOBA AI	Yes, very often the action space is growing during the game as new actions become available; it is slightly higher than for team sports at the start, with a low number of agent-specific actions.
RTS AI	Action space is dynamically growing, but this is less emphasized and not available for all unit types; movement action space is quite limited, and certain actions such as shooting are often done automatically.

Table 5. **Movements/state spaces** property. How free is the AI in controlling movement? What additional data is needed to make up the state space?

Team sports AI	In principle, agents can move over the whole field, but roles strongly restrict movement space in most situations; additional restrictions are imposed by other factors, such as physical conditions and penalties of the players.
MOBA AI	In principle, agents move freely, but in most situations (especially at the beginning) the movements are strongly restricted; possible movements are also affected by many additional properties, such as physical state and obtained items.
RTS AI	Free movement, but rarely used in practice (e.g., for scouting), since units have to stay together for successful combat; reasoning is mostly done on on the level of squads rather than individual units.

7   THE CHALLENGES OF SPORTS AI

One of the goals of AI competitions is to develop AI methods, applicable to a wider domain or at least to a wider set of subtasks in the given domain. As expressed by Togelius [33], *"we should try to counteract the tendency of competition participants to overfit their solution to particular problems and problem parameters, so as to make the results of competitions more generally valid."* This is a relatively recent approach for setting up competitions: it comes with the rise of *artificial general intelligence* (AGI), which requires a much higher level of autonomy than most AI approaches currently have. A recent example of a step into this direction is the Obstacle Tower Challenge [11] with its procedurally generated, diverse levels that require high generalization skills.

Table 6. **Rhythm/independent episodes** property. Is it possible to change the strategy easily (re-plan requirement)?

Team sports AI	Most sport games have a restart behavior: game state is relatively independent of what happened in the last episode.
MOBA AI	Frequent restarts in the beginning, towers have similar meaning as goals in soccer; the more successful party (with more kills) gains advantage over time.
RTS AI	No restart behavior: once one team has strategic dominance, it is hard to overturn the situation.

Likewise, team sport games can serve as a good vehicle for researching a diverse set of AI-related challenges, including the following:

(1) **What can coaches and athletes learn from AI and vice versa?** One can easily observe that the strategies of the top RoboCup 2D Simulation League teams are substantially different from the strategies of teams in real soccer. Despite the specific goals of RoboCup and their implications for the game (see Section 3), it is still unclear which parts of the setup can be responsible for these differences. We also do not know who, in principle, plays computer soccer better — skillful human players or AI teams. We can compare AI-controlled characters in games like FIFA with their real-world prototypes, but there has been no comparison between the best real soccer teams and AI-controlled computer soccer teams. It is possible that AI systems can greatly benefit from the arrival of digitized real soccer team behavior data; and it is also possible that real-world teams could benefit from studying team tactics, exhibited by the best RoboCup teams.

(2) **Which facets does a good team sport AI integrate?** Sport games almost inevitably have to rely on some type of AI technology. A computer soccer match without AI would require simultaneous presence of exactly 22 participants plus referees. Thus, the role of AI in this setup is to be both an opponent and a teammate, and to provide smooth gaming experience and "suspension of disbelief" for human players. Environments like RoboCup set a straightforward goal for an AI system: it has to win as many matches as possible. However, as AI methods mature and become more skillful, like in chess or fighting games, we will need another benchmark, emphasizing the *user enjoyment* facet, crucial for a *game* system. Player appeal is much harder to evaluate — the goal is elusive, and the benchmarks are subjective, but one cannot just ignore the existence and importance of this factor.

(3) **What is emergent and stable team behavior?** While dealing with team sports, one has to clarify what constitutes team behavior, and what are the characteristics of successful teams. It is possible that team behavior can be defined in terms of goal-driven decision-making, where team goals (e.g., scoring) take precedence over individual goals (e.g., showing off). However, real soccer teams possess other important traits, such as adaptability to opponent counter-actions, efficient repetition of the same successful patterns or adjusting strategies on the go. Professional players are able to predict the actions of their teammates and opponents (and quickly adapt their behavior if these predictions turn out to be wrong), as well as perform clearly identifiable elements of team strategy. Teams can quickly regroup after an unsuccessful attack and readjust formations in case of player injury or removal. Thus, team behavior is a complex interplay of individual and group tactics, which makes it an interesting challenge for an AI system with high potential impact.

8 DISCUSSION

Intelligent team behavior is an interesting problem of AI research, having great practical significance. Currently there are no established testbeds for benchmarking team AI behavior, and we propose that sport games similar to soccer, ice hockey or basketball have a potential to become such a testbed. Arguably, alternative options could include game genres like RTS and MOBA. According to our comparison (see Tables 1-6), MOBA games are especially close to sport games in several aspects. However, there are notable differences, too.

While successful strategies in RTS and especially MOBA games require team coordination, arguably, "team behavior" is not the most complex skill to master in these genres. RTS and MOBA are specifically designed to be enjoyable for all participants. Thus, every player typically controls a "hero", being in charge of their own group of subordinate units. Complex team-based strategies might require some participants to sacrifice their squad for the benefit of the team and/or concentrate on unpleasant menial tasks to support the frontline actually fighting the enemy. However, such patterns are not very common in real game sessions thanks to a carefully constructed design of game economy and team structure.

Team sport games have grown naturally around the concept of team-vs-team competition, and their rules are a product of long evolution, aimed at keeping the game interesting for both participants and spectators. While in each sport there is a number of superstar athletes showing exceptional abilities, generally, sport teams are comprised of people possessing comparable skills (in any case, there is no such "unit diversity" as in RTS), and game elements like 'resource mining' or 'research and construction' are absent. Thus, one of the ways to make the game complex enough to keep people's interest is to make room for diverse and non-trivial team strategies. This might mean that some team members have indeed to perform activities that can be perceived as "less exciting", like goalkeepers in soccer, who normally have no chances to score a goal. However, their somewhat auxiliary team role is compensated by the possibility to exhibit the personal mastery required to perform their role well.

Therefore, we believe that sport games may be at least as challenging as RTS/MOBA games in terms of developing efficient team strategies. Moreover, it is easier to start an AI research project with something relatively simple like two-player-team beach volleyball, continue with five-a-side football or futsal, and then proceed to a more complicated game like soccer or American football. The problem of AI creation is also limited in this case with a relatively uniform set of tasks, dealing with various aspects of team behavior, while in case of RTS/MOBA, a successful AI system has to address diverse issues like resource management, unit choosing, complex hierarchical goal planning, and so on.

In terms of the sheer number of game elements, sport games are much simpler than RTS/MOBA, and thus their game engines should be easier to develop and easier to communicate with. A considerable challenge of modern sport games development lies in the need to implement smooth and complex animated sequences for a diverse set of onscreen actions. However, this task is not relevant for AI research, and thus can be greatly simplified. If we look at existing team sport-like systems actually used for AI projects, such as RoboCup 2D Simulation platform, WeBots[4] (used in recent AI World Cup competitions) or MuJoCo Soccer Environment[5] (serving as a platform for DeepMind's experiments in team AI), we observe that all of them downplay animation and player contact, and instead focus on movement and passing behavior.

We must recognize, however, that there is no agreement on what constitutes "core" game elements that must be somehow represented in a computer simulation. For example, in soccer-like games

---

[4]https://www.cyberbotics.com/
[5]https://deepmind.com/research/open-source/mujoco-soccer-environment

we see environments that implement or ignore elements like physical contact/tackling, overhead passes and the offside rule. Thus, it is difficult to assert which particular elements are crucial for "interesting" team behavior (though, the rules of actual sports should serve as a reasonable approximation, and, probably, should not be ignored without good reasons).

Since sport games represent real sports, serious players have more elaborate expectations about AI, especially as other game aspects reach higher levels of realism. The AI-controlled opponents have to be believable, and the teammates have to be reasonably creative and supportive. It is likely that the growing availability of actual player tracking data can provide valuable insights for reaching these goals.

We can also note that the world of a typical sport game can hardly be characterized as "rich" — there are no fantasy-themed landscapes, intricate dungeons or exciting storylines. As there is not plenty of "decorative" game elements to "distract from" poor AI technology, it has to perform on par with other core elements, such as animations and physics. Maybe even more than in other game genres, some AI individuality is expected, e.g. in a football video game, where Messi and Ronaldo shall not only visually look like their real-world counterparts, but also somehow behave similar to them.

Another very interesting aspect of AI for team sport games is related to the diversity of possible goals: take a role of a teammate/opponent/referee; be a superstar forward player or a supportive defender; be strong and efficient, or be fun and inventive. Since the rise of really strong AI systems for a variety of games, it can be expected that these aspects of AI development would get more attention. As a firmly established activity, sport games can be good environments for investigating such goals, as we already know much about motivation of spectators and athletes, principles of good refereeing, and understand the general psychological environment surrounding sports competitions.

9 CONCLUSION

This paper has argued that team sports occupy a special, possibly unique, niche in the AI development landscape, and thus deserve attention on their own, rather than within the context of other multi-agent environments, such as RTS or MOBA games. Sport video games are virtual representations of real-world competitions, and thus hold a borderline position between the real world and the virtual world. On the one hand, they have to capture the essence of real-life events and represent them faithfully to be appealing for sports fans. On the other hand, they have to be fun to play and accessible for gamers, who want to be immersed in a make-believe world, where they can play the roles of successful top-class athletes.

A general trend in the sport games genre is convergence between the game and the reality [27], which in practice means more realistic graphics, environments, and AI systems. Good AI is one of the cornerstones of a high-quality team sport game; it is hard to imagine, for example, a soccer-like game set up in a people-only online multiplayer mode, and played without AI.

Currently, the advancements in AI technologies extend the list of games considered "solved" in the sense that AI beats professional human players. In practice, this means that the focus of game AI research can shift to secondary challenges, aimed at maximizing entertainment value of game products. Sport games are especially good for studying such emotional-driven goals of AI: the pool of popular sports events is stable, and we know much about motivation and enjoyment of both participants and fans.

One may argue that the issues related to subjective user enjoyment are hard to deal with: there are no easy ways known to identify AI-related sources of user enjoyment, to implement AI that maximizes fun, nor to evaluate obtained results. However, dealing with *sport game* AI systems, it is impossible to ignore these factors, especially in the light of their growing importance. It is,

therefore, both reasonable and necessary to anticipate this trend, by focusing more research effort towards defining an appropriate and long-lasting AI benchmarking environment.

ACKNOWLEDGMENTS

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] Hidehisa Akiyama and Tomoharu Nakashima. Helios base: An open source package for the RoboCup soccer 2D simulation. In Sven Behnke, Manuela Veloso, Arnoud Visser, and Rong Xiong, editors, *RoboCup 2013: Robot World Cup XVII*, pages 528–535, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[2] Ryan Beal, Timothy J. Norman, and Sarvapali D. Ramchurn. Artificial intelligence for team sports: a survey. *The Knowledge Engineering Review*, 34:e28, 2019.

[3] G. Bradley and C. Toye. *Playing Soccer the Professional Way*. Harper & Row, 1973.

[4] Collins. Collins english dictionary complete and unabridged (13th ed.), 2018.

[5] Abel Correia and Sandra Esteves. An exploratory study of spectators' motivation in football. *International Journal of Sport Management and Marketing*, 2(5-6):572–590, 2007.

[6] Kevin Dill. What is game AI? In Steve Rabin, editor, *Game AI pro*, pages 3–10. A K Peters/CRC Press, 2013.

[7] Alexander Ferrein and Gerald Steinbauer. 20 years of RoboCup. *KI-Künstliche Intelligenz*, 30(3-4):225–232, 2016.

[8] FIFA. Goalkeepers are not above the law, 1997.

[9] Gandolfi, G. (Ed.). *NBA Coaches Playbook: Techniques, Tactics, and Teaching Points*. Human Kinetics, 2008.

[10] Johnston, M. and Walter, R. *Hockey Plays and Strategies, 2nd Ed.* Human Kinetics, 2018.

[11] Arthur Juliani, Ahmed Khalifa, Vincent-Pierre Berges, Jonathan Harper, Hunter Henry, Adam Crespi, Julian Togelius, and Danny Lange. Obstacle tower: A generalization challenge in vision, control, and planning. *CoRR*, abs/1902.01378, 2019.

[12] Graham Kendall and Liam J.A. Lenten. When sports rules go awry. *European Journal of Operational Research*, 257(2):377 – 394, 2017.

[13] Christina Koch and Markus Tilp. Beach volleyball techniques and tactics: A comparison of male and female playing characteristics. *Kinesiology*, 41:52–59, 06 2009.

[14] Patrick MacAlpine and Peter Stone. UT Austin Villa RoboCup 3D simulation base code release. In Sven Behnke, Daniel D. Lee, Sanem Sariel, and Raymond Sheh, editors, *RoboCup 2016: Robot Soccer World Cup XX*, Lecture Notes in Artificial Intelligence, pages 135–43. Springer Verlag, Berlin, 2017.

[15] John McCarthy. Chess as the drosophila of AI. In *Computers, chess, and cognition*, pages 227–237. Springer, 1990.

[16] Olivia Michael and Oliver Obst. Betarun soccer simulation league team: Variety, complexity, and learning. *arXiv preprint arXiv:1703.04115*, 2017.

[17] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing Atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[18] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, February 2015.

[19] Inseok Oh, Seungeun Rho, Sangbin Moon, Seongho Son, Hyoil Lee, and Jinyun Chung. Creating pro-level AI for real-time fighting game with deep reinforcement learning. *arXiv preprint arXiv:1904.03821*, 2019.

[20] Santiago Ontañón, Gabriel Synnaeve, Alberto Uriarte, Florian Richoux, David Churchill, and Mike Preuss. A survey of real-time strategy game AI research and competition in StarCraft. *IEEE Trans. Comput. Intellig. and AI in Games*, 5(4):293–311, 2013.

[21] Mikhail Prokopenko and Peter Wang. Disruptive innovations in RoboCup 2D soccer simulation league: from cyberoos'98 to gliders2016. *CoRR*, abs/1612.00947, 2016.

[22] Mikhail Prokopenko, Peter Wang, Sebastian Marian, Aijun Bai, Xiao Li, and Xiaoping Chen. RoboCup 2D soccer simulation league: Evaluation challenges. In *RoboCup 2017: Robot World Cup XXI [Nagoya, Japan, July 27-31, 2017].*, pages 325–337, 2017.

[23] RoboCup. RoboCup Federation.

[24] RoboCup. RoboCup federation: Objective.

[25] RoboCup. RoboCup soccer simulation.

[26] Tara K. Scanlan, Paul J. Carpenter, Marci Lobel, and Jeffery P. Simons. Sources of enjoyment for youth sport athletes. *Pediatric exercise science*, 5(3):275–285, 1993.

[27] Miguel Sicart. A tale of two games: football and FIFA 12. In *Sports Videogames*, pages 40–57. Routledge, 2013.

[28] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, January 2016.

[29] P. Sweetser, D. Johnson, J. Sweetser, and J. Wiles. Creating engaging artificial characters for games. In *Proceedings of International Conference on Entertainment Computing*, pages 1–8, 2003.

[30] M. Swiechowski. Game AI competitions: Motivation for the imitation game-playing competition. In *2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, pages 155–160, 2020.

[31] Tahtouh, Toni Faouzi. *Volleyball: Techniques and Tactics*. Lulu Publishing Services, 2017.

[32] Leigh Thompson. *Making the Team: A Guide for Managers*. Pearson Education, 5th edition, 2014.

[33] Julian Togelius. How to run a successful game-based AI competition. *IEEE Transactions on Computational Intelligence and AI in Games*, 8(1):95–100, 2016.

[34] Yannick Verhoeven and Mike Preuss. On the potential of Rocket League for driving team AI development. In *Proceedings of IEEE SSCI 2020*, pages nn–mm, 2020.

[35] Oriol Vinyals, Igor Babuschkin, Junyoung Chung, Michael Mathieu, Max Jaderberg, Wojciech M. Czarnecki, Andrew Dudzik, Aja Huang, Petko Georgiev, Richard Powell, Timo Ewalds, Dan Horgan, Manuel Kroiss, Ivo Danihelka, John Agapiou, Junhyuk Oh, Valentin Dalibard, David Choi, Laurent Sifre, Yury Sulsky, Sasha Vezhnevets, James Molloy, Trevor Cai, David Budden, Tom Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Toby Pohlen, Yuhuai Wu, Dani Yogatama, Julia Cohen, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Chris Apps, Koray Kavukcuoglu, Demis Hassabis, and David Silver. AlphaStar: Mastering the Real-Time Strategy Game StarCraft II. https://deepmind.com/blog/alphastar-mastering-real-time-strategy-game-starcraft-ii/, 2019.

[36] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy P. Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. Starcraft II: A new challenge for reinforcement learning. *CoRR*, abs/1708.04782, 2017.

[37] Wikipedia contributors. Team sport — Wikipedia, the free encyclopedia, 2019.

[38] J. Wilson. *Inverting the Pyramid: The History of Football Tactics*. Orion, 2010.

[39] Wilson, J. Why is the modern offside law a work of genius?, 2010.

[40] Zauli, Alessandro. *Soccer: Modern Tactics*. Reedswain, 2011.