Game Theory in Soccer

Daniel Martin

May 2021

1 Introduction

Preface

We have recently entered the age of advanced analytics in sports. Professional teams of all sports are hiring analysts to evaluate players and to enhance coaching tactics. Naturally, these tactics are constantly changing and every team wants to have a dominant strategy. The implementation of different strategies in sports has been an interest of mine from a very young age. I grew up playing just about every sport there is. A lot of my time in high school was spent playing soccer, hockey, and golf in each of the school's sports seasons. Currently, I play hockey for The College of New Jersey.

Overview

My research will be based on studying different tactics, particularly different types of formations in a simulated game called DMA Soccer (defense, midfield, attack). Before I get into the game, I will explain definitions in both fields of probability and game theory. Then we will get into the game as in depth as possible and see what happens when different tactics get altered. My goal is to find an equilibrium of the strategies. I will be looking to see if there are any strategies that are better than comparative strategies. Hopefully by looking at all possible strategies and comparing them we will find an equilibrium of one or a couple strategies that are better than the rest.

2 Background

Definitions

Before DMA soccer is delved into, there are some helpful probability and game theory definitions that contextualize DMA soccer. We will start with some fundamental probability theorems. They each involve finding the probability of certain events occurring, where events are possible outcomes of the whole sample space.

1. Conditional Probability ([3] Definition 3.1)

   Given events A and B,

$$\Pr(A|B) = \frac{\Pr(AB)}{\Pr(B)}. \tag{1}$$

2. Bayes Theorem ([3] Theorem 3.3)

   Given event A and disjoint events B_1, B_2, ... , B_n

$$\Pr(B_i|A) = \frac{\Pr(A|B_i)\Pr(B_i)}{\Pr(A)}. \tag{2}$$

   For the simulations, I am going to look heavily at goal distributions and win loss distributions. Conditional probability will be used constantly to determine different probabilities. To calculate the multiple conditional probabilities, I will use Bayes Theorem.

3. Strictly Dominant Strategy [4]

   Let a and b be strategies in a game. We say that a strictly dominates b if the payoff for using strategy a is higher than the payoff for using strategy b. We will denote this as $A(a) > B(b)$ for teams A and B using strategies a and b, respectively. In the context of DMA soccer, the strategies will be the formations. This will be explained further when the rules of the game are explained, but in mathematical terms this can be thought of as the distribution of players in the different sections of the field.

4. Weakly Dominant Strategy

   Let a and b be strategies in a game. We say that a weakly dominates b if the payoff for using strategy a is higher than or equal to the payoff for using strategy b. We will denote this as $A(a) \geq B(b)$ for teams A and B using strategies a and b, respectively. Any strategy that is strictly dominant to another strategy will also be weakly dominant.

5. Nash Equilibrium

   The Nash Equilibrium is the scenario where both sides of a game choose their best strategy, regardless of the opponent's strategy. The Nash Equilibrium may not always exist if at least one side does not have a clear best strategy. We will denote this as $N(A, B) = (a, b)$, where Team A's best strategy is a and Team B's best strategy is b.

Applications of Definitions

The game theory definitions will be very important when we talk about different strategies in DMA Soccer. To explain them thoroughly, we will show an economics example. Consider two companies, Company A and Company B. The table below shows all the possible outcomes for the two companies based on their decision to either advertise, or to not advertise. The ordered pairs are in millions of dollars, where the ordered pair (c, d) represents the profit. The profit of Company A is represented by c and the profit of Company B is represented by d.

	B advertises	B doesn't advertise
A advertises	3, 3	6, 2
A doesn't advertise	2, 6	5, 5

Figure 1: Payoff Table [2] (Example 2.1)

We will let the strategy to advertise be strategy a and the strategy to not advertise be strategy b. In the chart above, we see the payoffs in ordered pairs where the first number is the payout for Company A and the second number is the payout for Company B. For example, if Company A and Company B both advertise (strategy a), then each of their payoffs will be 3. Now consider only Company A's payouts. If Company A does not advertise, their payoff will either be 2 if B advertises or 5 if B does not advertise. If Company A does advertise, then their payoff will either be 3 if B advertises or 6 if B does not advertise. We can see that regardless of what Company B does, Company A should advertise since it's payoff will increase ($2 \to 3$ if B advertises or $5 \to 6$ if B does not advertise). Then for Company A, the strategy of advertising is strictly dominant to the strategy of not advertising ($A(a) > A(b)$). Since 2 is strictly less than 3 and 5 is strictly less than 6, we use the stronger implication with the term strictly dominant instead of the term weakly dominant, although it would still apply. If we apply the same concepts to Company B, we will see that $B(a) > B(b)$. Since both Company A and Company B have strictly dominant strategies, a Nash Equilibrium will exist at $N(A, B) = (a, a)$.

3 The Game of DMA Soccer

The Game

My plan is to study the DMA soccer simulations and see what theories I can make based on results from these simulations. DMA soccer is a game where each team has a certain amount of players that they can put in any of three zones. These players cannot move between the zones. A ball starts out in the middle zone with each player in that zone having an equal chance to get the ball and kick the ball forward into the next zone, where the same scenario will occur with the players in that zone. The goal is to get your team to kick the ball forward through the last zone which will give your team a goal. The game goes on for a given number of rounds, and each time the ball is thrown to a different zone will be half of a round. We will be looking to determine what strategies are dominant.

Visualization

Below is a visualization of DMA soccer with teams of 5.

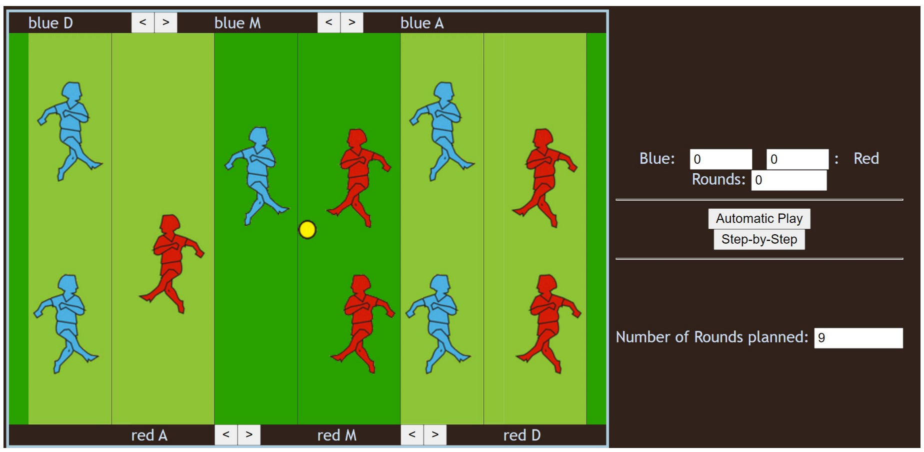

Figure 2: Visualization [2](Applet DMAstatic)

Calculations for One Round

Below is a figure demonstrating how Bayes Theorem will be used to calculate the probabilities for the goal distribution. In this figure, the pair of letters stands for the number of players in that specific section. The first letter stands for the position: A for attack, M for midfield, and D for defense. The second letter stands for the different teams: we let A be the blue team, and B be the red team as depicted in the picture above. (The left team will be blue, and the right team will be red).

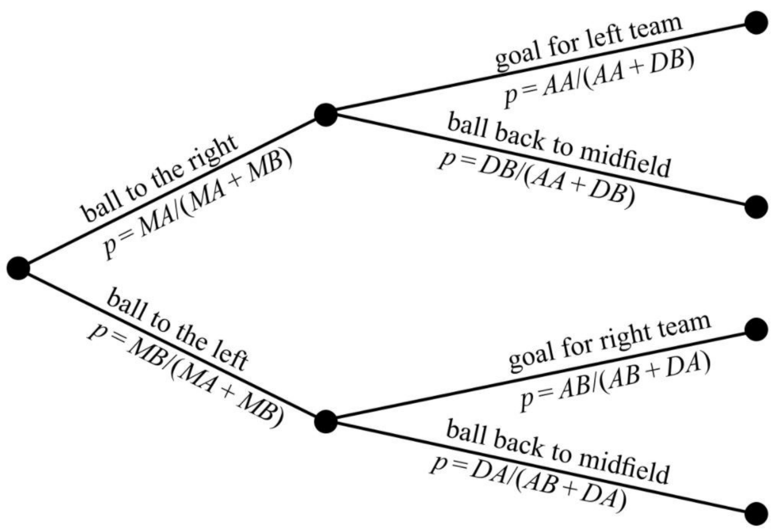

Figure 3: Probability Tree for One Round [2] (Figure 14.1)

Example of Calculation

Consider the formations in Figure 3. The left team (Blue) has a 2-1-2 formation (2 defenders, 1 midfielder, and 2 attackers), while the right team (Red) has 2-2-1 formation. The ball starts in the middle so we look at the midfielders first, and start from the left side of the probability tree. Consider the probability that the ball goes to the right ($\Pr(R1)$). This will mean that the left team will win the ball forward. The necessary terms for the calculation are MA and MB. Since we know the formations, we know both of these terms. $MA = 1$ since Blue has one player in the midfield area. $MB = 2$ since Red has two players in the midfield area. So $\Pr(R1) = \frac{1}{1+2} = \frac{1}{3}$. Conversely, $\Pr(L1) = \frac{2}{1+2} = \frac{2}{3}$. Now assume that the ball went to the right. There are two more options: Blue will win the ball forward (right) for a goal, or Red will win the ball forward (left) and the ball will go back to midfield. The two new terms we need for the calculation are $AA = 2$ and $DB = 2$. So $\Pr(R2|R1) = \frac{2}{2+2} = \frac{1}{2}$. Conversely, $\Pr(L2|R1) = \frac{1}{2}$. We can finish the tree diagram by calculating the two options under the assumption that the ball goes left from midfield. So $\Pr(R2|L1) = \frac{2}{3}$ and $\Pr(L2|L1) = \frac{1}{3}$. Now that we have the probability tree figured out, we can apply Bayes Theorem to calculate other probabilities such as the probability that the blue team scores when the ball starts in the middle. The probability we are looking for is $\Pr(R1 \text{ and } R2)$ because we want the ball to be kicked to the right twice. From Bayes, $\Pr(R1 \text{ and } R2) = \Pr(R2|R1)\Pr(R1)$. Substituting in for $\Pr(R2|R1)$ and $\Pr(R1)$ gives us $\Pr(R1 \text{ and } R2) = \frac{1}{2}\frac{1}{3} = \frac{1}{6}$. This tree diagram shows just for one round and one combination of formations, so there will be many of these calculations.

Expected Goals

The calculations above show all of the probabilities for one round, but we need something more to compare strategies. To find dominant strategies, we need to define what the payoff will be for DMA soccer. In most economics examples we would expect payoff to be something like profit. In DMA soccer, expected goals is the payoff. I will define the expected goals for Team A in one round as the probability that Team A scores a goal in that round. In equation form, this is

$$E(A) = \Pr(A \text{ scores a goal}) = \frac{MA}{MA + MB} \frac{AA}{AA + DB}.$$

When strategies are compared, the expected goal differential between two teams will be considered. The expected goal differential for Team A and Team B is the difference between the expected goals for Team A and the expected goals of Team B. In equation form, this is

$$E(X) = E(A) - E(B).$$

Example of Calculation

The calculations for expected goals are similar to those of the probability tree above. Given the same formations as above (2-1-2 for Blue, and 2-2-1 for Red), $E(Blue) = \Pr(R1 \text{ and } R2) = \frac{1}{6}$ from above. We can use the formula to find the expected goals for red. This is

$$E(Red) = \Pr(L1 \text{ and } L2) = \frac{MB}{MB + MA} \frac{AB}{AB + DA} = \frac{2}{(2+1)} \frac{1}{(1+2)} = \frac{2}{9}.$$

Now that we have $E(Blue)$ and $E(Red)$, the expected goal differential can be calculated. $E(X) = E(Blue) - E(Red) = \frac{1}{6} - \frac{2}{9} = \frac{-1}{18}$. All of these calculations can be done using Microsoft Excel.

4 Results

5 Player Strategies

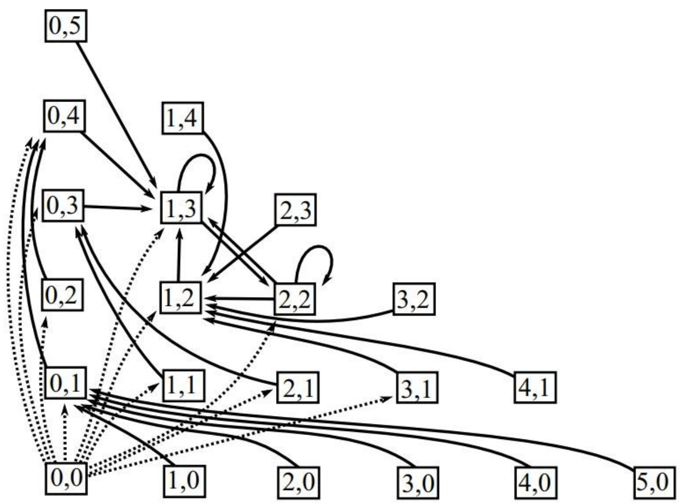

Figure 4: Best Response Diagraph for goal difference [2]

Figure 4 is a diagraph showing the relationships between all possible strategies for 5 players per team. Each ordered pair (A, B) stands for one formation. A is the number of defensive players, B is the number of midfield players, and the number of attacking players is $5 - (A + B)$ since there are 5 total players. The arrows point to the strategy that is weakly dominant. For example, (0, 5) is weakly dominated by (1, 3). This means that the expected goals for the 1-3-1 formation is higher than the expected goals for the 0-5-0 formation when they are matched up against each other. The goal of the research is to find out what strategies are the best. Strictly looking at expected goal differential as the payout, there is an equilibrium at the 1-3-1 and 2-2-1 strategies. It is interesting how the 2-2-1 strategy is part of the equilibrium, but is not symmetric like the 1-3-1 strategy is. The 2-2-1 strategy has one more defensive player than attacking player, so we can consider it a relatively defensive strategy. Before doing the calculations, I expected the 1-3-1 strategy to be part of the equilibrium. My initial thoughts were that the midfield would have to have at least 2 players since the ball starts there and is there most often. Each round will have the midfielders for each team battle, specifically in the first half of the round. If the midfielders win the ball forward, then the defensive players are not used in that round whether the attackers end up scoring or losing the ball back to the midfield. I was not surprised to see that strategies that had at least one player in each zone fared much better than strategies that had at least one zone with no players. Because of the way the strategies are being considered with regards to expected goal differential, there is an assumption being made that

DMA soccer is a zero-sum game. DMA soccer can be viewed as a non-zero-sum game though.

Zero-Sum Games

A zero-sum game is a game where the expected payoff between both strategies sums to zero. As an example, Major League Baseball is a zero-sum game. Although they technically do not have a points system for wins, losses, and ties, they do have it implicitly. The way teams are compared is by a statistic called "games back". The way games back is calculated is by giving a win 0.5 points and a loss -0.5 points. Since there are no ties and each game will have one winner and one loser, the points sum to zero $(0.5 + (-0.5) = 0)$. For example, let's say Team A has played 5 games and has a record of 3 wins and 2 losses. Let's say Team B has played 3 games and has a record of 1 win and 2 losses. Then Team B would be 1 game back of Team A since $([(0.5) * 3 + (-0.5) * 2] - [(0.5) * 1 + (-0.5) * 2] = [1.5 + (-1)] - [0.5 + (-1)] = 1)$. Not all sports are zero-sum games though, and we can make DMA soccer a non-zero-sum game.

Non-Zero-Sum Games

Other sports leagues such as the National Hockey League (hockey) and the English Premier League (soccer) are not zero-sum games. The National Hockey League (NHL) has a point system awarding 2 points for a win, 1 point for a loss past regulation time (in overtime or a shootout), and 0 points for a loss. Some games will have a total sum of 2 points if the game ends in regulation time, and some games will have a total sum of 3 points if the game ends in overtime or a shootout. Similarly, the English Premier League (EPL) has a point system awarding 3 points for a win, 1 point for a tie, and 0 points for a loss. Again, the sum of points for each game changes between 3, if there is a winner and a loser, and 2, if both teams tie. To see how these point systems affect the equilibrium for strategies in DMA soccer, we will look at 3-round games with 5 players per team.

3-Round Game

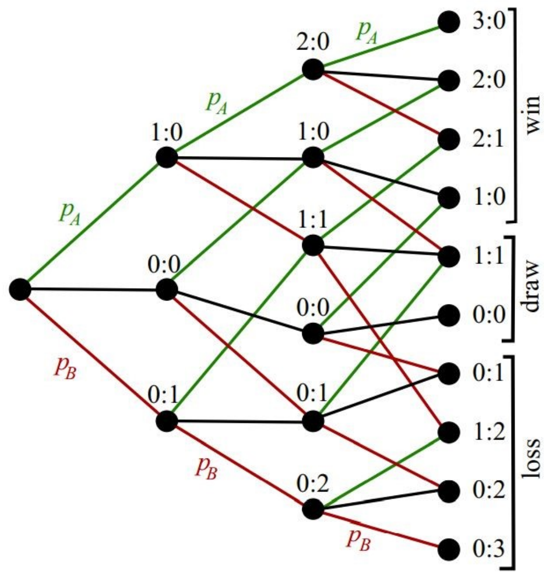

Figure 5: Probability Tree for 3-Rounds of DMA Soccer [2]

Figure 5 shows all of the possible scores in a 3-round game, and which outcomes result in a win, draw, or loss. Calculating these probabilities is the same process as the one round game and adding up the possibilities for the possible scores.

Example

Some of the score outcomes will not be complex. For example, Pr(3:0 win) = Pr(goal and goal and goal). Sticking with the 2-1-2 and 2-2-1 formations for the previous calculations,

$$\text{Pr(3:0 win)} = \text{Pr(R1 and R2)} \times \text{Pr(R1 and R2)} \times \text{Pr(R1 and R2)} = (\frac{1}{6})^3 = \frac{1}{216}.$$

The only possible way for this outcome to happen is for a goal in each round. When looking at something more complex such as Pr(2:0 win), there are 3 possible ways to get to that score. 2 rounds will have a goal and 1 round will not. So the 3 possible outcomes are

Pr(goal and goal and no goal),

Pr(goal and no goal and goal),

and

$$\Pr(\text{no goal and goal and goal}).$$

Since these 3 probabilities are all equal,

$$\Pr(2{:}0 \text{ win}) = 3\Pr(\text{goal and goal and no goal}) =$$

$$= 3\Pr(\text{R1 and R2})\Pr(\text{R1 and R2})[\Pr(\text{R1 and L2}) + \Pr(\text{L1 and R2})] =$$

$$= 3\frac{1}{6}\frac{1}{6}(\frac{1}{6} + \frac{2}{9}) = 3\frac{7}{648} = \frac{7}{216}.$$

There will only be 3 combinations each to get the winning scores of 2:1 and 1:0. All of these calculations have been done in excel.

Expected Points

Now that the probability for each strategy to win, tie, or lose has been calculated, points systems can be looked at in depth. Consider the points systems of 3-1 (3 points for a win, 1 point for a draw, and 0 points for a loss) and 3-2. Instead of looking at expected goals, expected points will be looked at. The formula for expected points for Team A is $E(P_A) = w*\Pr(W)+d*\Pr(D)+l*\Pr(L)$ where w, d, and l are the points for a win, draw, and loss respectively and $\Pr(W)$, $\Pr(D)$, and $\Pr(L)$ are the probabilities of a win, draw and loss respectively. Similar to how expected goal differential was used to compare strategies, expected point differential will now be used.

3-1 and 3-2 Point System Results

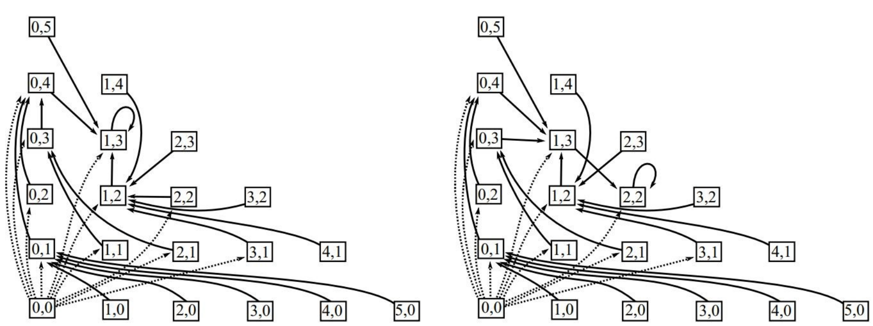

Figure 6: Best Response Diagraph for 3 rounds with 3-1 and 3-2 Point Systems [2]

The graphs are formatted the same as the one for expected goal differential. The graph on the left is for the 3-1 point system and the right is for the 3-2 point system. In the 3-1 point system, the equilibrium is now only the 1-3-1 formation. In the 3-2 point system, only the 2-2-1 formation is part of the equilibrium. Without changing the rules of the game, the point system can directly effect team's strategies. In the 3-1 system, strategies that are more defensive, like the 2-2-1 strategy, are not as effective. Settling for a tie instead of going for a win is not as attractive as it once was in the zero sum game. Conversely, in the 3-2 point system, relatively aggressive strategies like the 1-3-1 formation are no longer as effective. In this case, settling for a tie instead of going for the win seems to be the right decision to maximize expected points.

5 Results for 6 Players per team

Similar to what was done for the results for 5 players per team, response diagraphs were made both for expected goal differential and expected points.

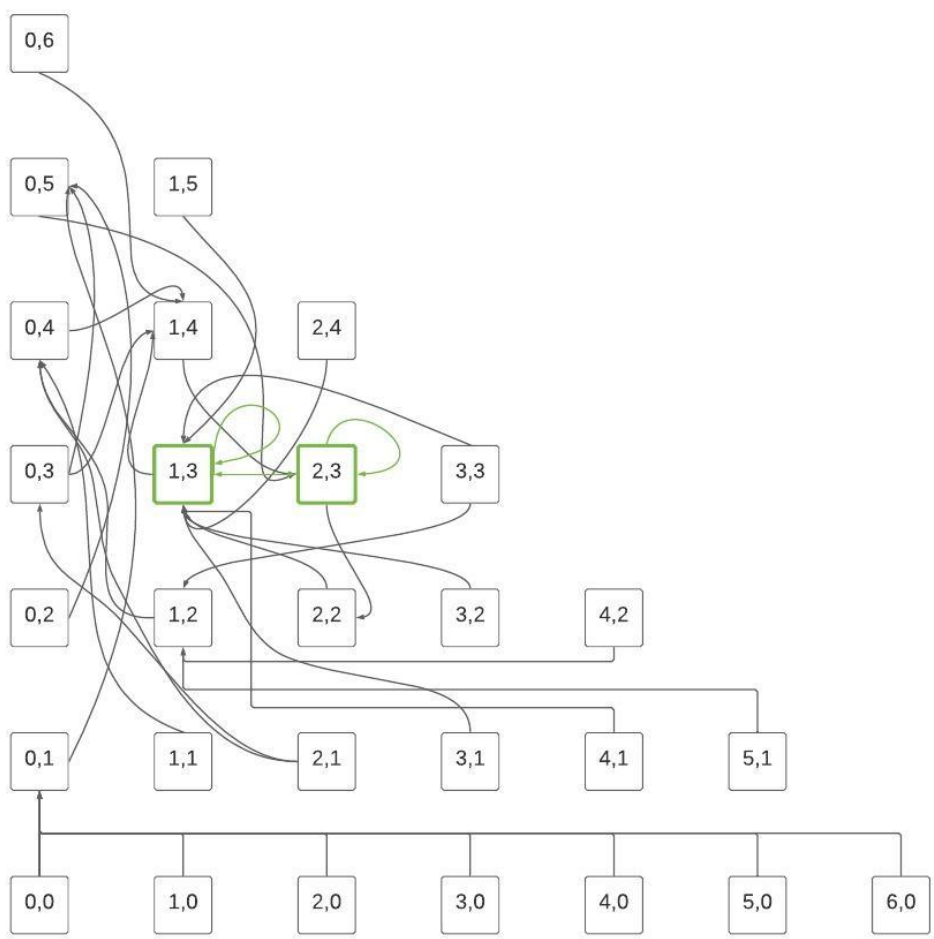

Figure 7: Response Diagraph for 6 Players by Expected Goal Difference

Figure 7 is formatted and abbreviated in the same way that Figure 4 is. We can see the response diagraph is very similar to the one in Figure 4. Instead of 5 players per team in each strategy, now we have 6. So for each strategy (A, B), the number of attacking players is $6-(A+B)$. Here we can see that the equilibrium is composed of two strategies: the 1-3-2 and the 2-3-1 strategies.

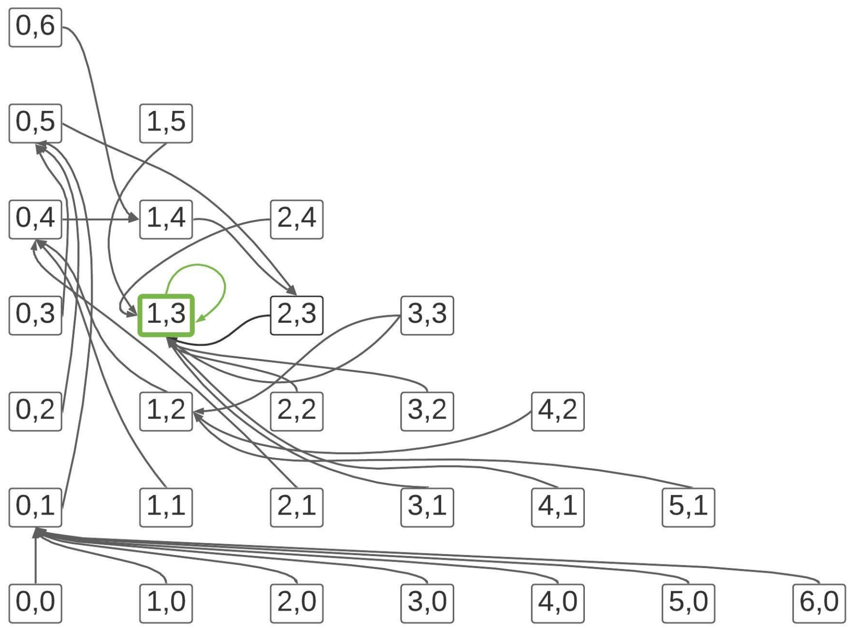

Figure 8: Response Diagram for 6 Players with a 3-1 Point System

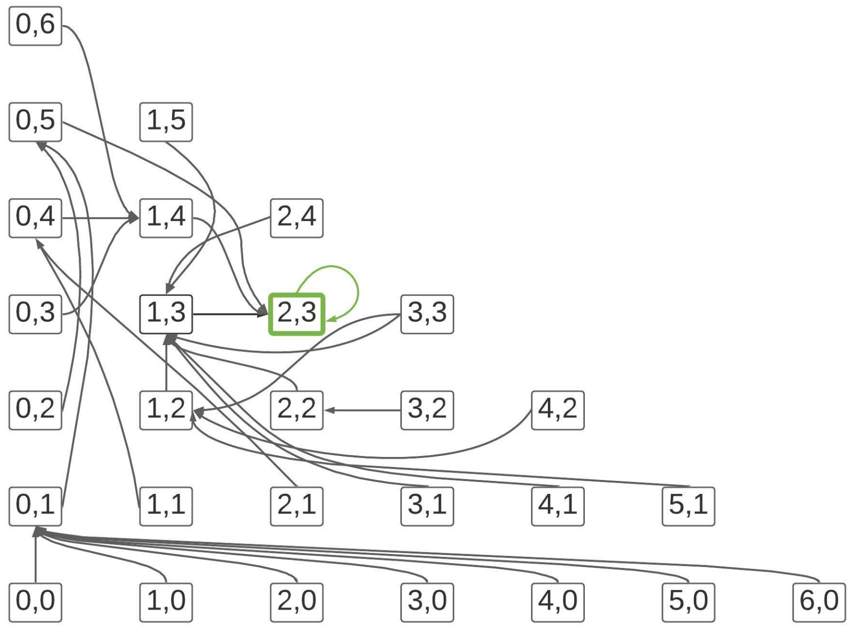

Figure 9: Response Diagraph for 6 Players with a 3-2 Point System

Figures 8 and 9 show the best response diagraph for 3 rounds with the different point systems. There are similar trends that were also seen in Figure 6. Defensive strategies are more effective in the 3-2 point system and offensive strategies are more effective in the 3-1 point system. The 3-1 point system has a single equilibrium at the 1-3-2 strategy where offense is more rewarding. The 3-2 point system has a single equilibrium at the 2-3-1 strategy where defense is more rewarding. We would expect these trends to continue if we added more players per team.

6 Conclusion

While DMA soccer has some restrictions that sports do not have, there are some ideas from DMA soccer that are present in sports. Specifically, point systems are an effective way for leagues to control scoring without changing the rules of the game. Leagues that want teams to increase goal scoring can increase the discrepancy between a win and tie so that teams have to respond more aggressively. Earlier in the paper, the NHL's 2-1 point system was discussed. A common complaint for the league is that teams too often stop playing offense in the last period of the game in order to hold their lead or play to get a guaranteed point (if the game is tied after regulation time). Below is a graphic showing how teams stop attacking when they are up one goal. The y-axis is percent of expected goals for relative to the average expected goals for. Each color represents the order in which the goals are scored. For example, "ahh" represents the away team scoring first and the home team scoring the next two goals.

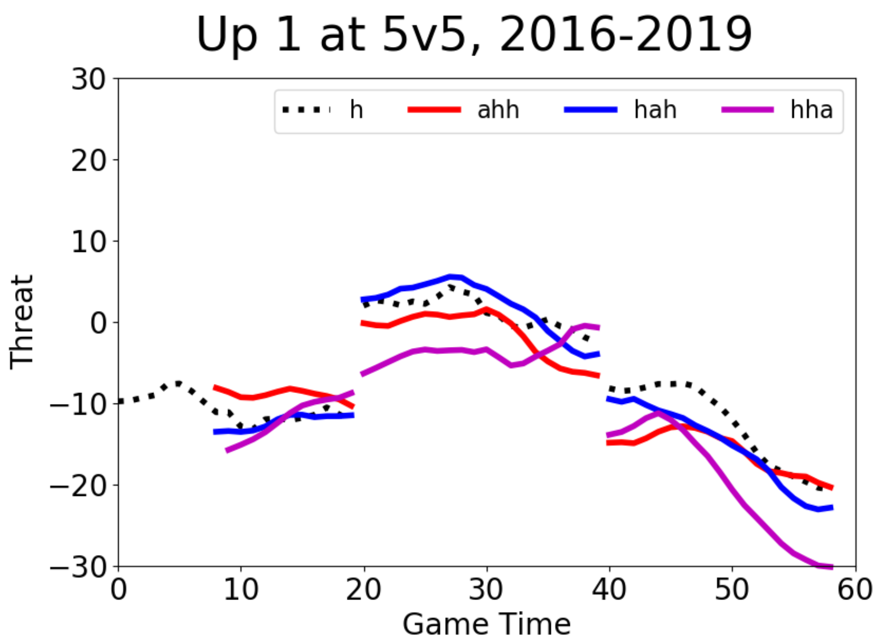

Figure 10: Expected Goals For Threat Over Time [1]

Figure 10 shows how teams tighten up defensively and stop trying to score. While the last few minutes in a one goal game are usually exciting, minutes 45 to 55 could use some more excitement. Presumably, a more aggressive point system would fix this. Another league that could be benefited by reviewing their point system is Major League Baseball. In the last few years the MLB implemented a rule for games that are tied after 9 innings. To end the games quicker, teams will start with a runner on second base to start the inning. There has been much criticism about this rule from baseball fans for changing the rules of the game. One solution to rectify this problem would be to add a non-zero-sum points system. They could mirror the NHL's point system, where there are no ties. They could also use a 3-1 point system, and make the 1 point a loss in extra innings. Another league that could benefit by recognizing the impact of points systems would be Major League Soccer (MLS). The MLS has a 3-1 point system for the regular season to determine standings. When teams play in the playoffs, this point system changes implicitly to a zero sum point system where there are no ties and the losing team is eliminated. It would be helpful if coaches recognize that strategies in the regular season may not be as effective in the playoffs, and the equilibrium is likely different. While the strategies in a stagnant game of DMA soccer may not transcend to different sports, the effect points systems have on strategies will.

References

[1] Micah Blake McCurdy, 2021.

[2] Erich Prisner, *Game theory through examples*, 1 ed., Mathematical Association of America, 2014.

[3] Richard Scheaffer and Linda Young, *Introduction to probability and its applications*, Cengage Learning, 2009.

[4] Bernhard von Stengel, *Game theory basics*, London School of Economics, 2011.