Using Markov Chains and Statistical Analysis to Model Intentional Fouling Situations in NCAA Division 1 Men's Basketball

Anirudh Sengupta (Author)

Ardrey Kell High School, Charlotte, North Carolina

North Carolina Summer Ventures in Science and Mathematics

East Carolina University

Dr. Susan Mattingly (Mentor)

Abstract

Intentionally fouling a team during their offensive possession is a strategy that basketball teams have employed for many years. Though popularized in the National Basketball Association (NBA), the technique has also made its way into Collegiate Basketball in recent years. The purpose of intentionally fouling is often to slow down opponent scoring and expose poor free-throw (FT) shooters. Despite previous studies using Markov Chains or analyzing intentional fouling in the NBA, there is a lack of research combining these methods with collegiate men's basketball rules, particularly the one-and-one bonus. This paper will examine the optimal times to start intentionally fouling when trailing, using Markov Chains and additional statistical analysis of data from the National Collegiate Athletic Association Division 1 Men's Basketball (NCAA D1 MBB). Additionally, this study will account for the Bonus and Double Bonus rules specific to NCAA Men's Basketball. Play-by-play data was compiled from over 4,000 games in the 2024-25 season to calculate statistics that would help identify the probabilities of different plays. Transition matrices were created to determine the expected points in a possession with and without fouls. By categorizing data based on the time remaining in the game, the expected points for all Division 1 teams in various situations were calculated, enabling the development of a strategy to identify the optimal time for a team to benefit from fouling. The specific examples of the Florida Gators, Alabama Crimson Tide, and App State Mountaineers demonstrate that the optimal time to start fouling changes based on the opposing team. Additionally, the matrices were used to calculate the maximum FT% of a player for which it would be optimal to foul, rather than to give up a possession. While situations change on a game-by-game basis, the graphs created in this study will provide a general idea of when teams are expected to score fewer points in free throw situations compared to offensive possessions.

Introduction

In the game of basketball, a foul is an infraction of the rules that is charged to a player or bench personnel and is penalized in various ways (NCAA, 2024). The most common penalty for a foul is to award the opposing team with Free-Throws (FTs), opportunities in which a player can score one point by an unhindered try for goal from within the FT semicircle and behind the FT line. During FTs, the game clock is paused and will resume once the FTs have ended. Though this rule was meant to deter fouling, it ended up creating a new defensive strategy to slow opponents and expose poor FT shooters. The most notable example of this was the "Hack-a-Shaq" method in the NBA, where teams on defense intentionally fouled poor FT shooters, such as Shaquille O'Neal (Young, 2021). Team A, by fouling poor shooters, would prevent Team B, their opponents, from gaining offensive momentum, while leaving more time on the game clock, meaning there are more scoring opportunities for themselves. This strategy has been heavily implemented in the NBA, as well as in NCAA Men's Basketball. Toward the end of a game, when a team is losing by a sizable margin, it is generally expected for that team to foul in order for them to have more offensive opportunities. When analyzing intentional fouling in college basketball, it is essential to consider the unique rules that come in NCAA Men's Basketball. The penalties for specific foul counts are listed below in Figure 1.

Number of Personal/Common Fouls in a Half	**Penalty**
1-6	No FTs awarded. The offensive team will receive an inbounds pass and continue possession.
7-9 (Bonus/One-and-One)	The fouled player will shoot one FT. If the FT is successful, the player gets an attempt to shoot a bonus FT. If the bonus FT is successful, the team will receive a second point. If either attempt is unsuccessful, the ball becomes live and the game clock will continue. This means that if the first FT is missed, the ball will become live and both teams will have the opportunity to gain possession.

10+ (Double Bonus)	The fouled player will shoot two FTs. If the second FT is missed, the ball will become live and both teams will have the opportunity to gain possession.

Figure 1: Table Describing Penalties for Different Free Throw Counts in NCAA Men's Basketball

An additional type of foul is a shooting foul, which occurs when a defensive player makes illegal contact with an offensive player while they are in the act of shooting. If a player is fouled during a 2-PT Attempt, they will receive 2 FTs, and if a player is fouled during a 3-PT Attempt, they will receive 3 FTs. For this study, shooting fouls will be ignored; however, FTs that result from 2-PT attempts could be considered the same as in a Double Bonus FT situation, because the offense is awarded two FTs.

Background

Intentional fouling in the NBA has been extensively researched, with numerous analysts and teams seeking to determine the optimal strategy for fouling. A study by Franklin H. J. Kenter (2015), presented at the MIT Sloan Sports Analytics Conference, analyzed NBA data from 2007 to 2014 to identify optimal plays based on point differential and time remaining in the game. The paper treated basketball as a combinatorial game, where Kenter analyzed the payoff of fouling in order to find the optimal strategy. A similar paper conducted by Julian Zapata-Hall (2025) focused on the last 30 seconds of NCAA Men's Basketball games. After analyzing data from 2009 to 2022, he developed a strategy that informed coaches whether they were better off fouling or defending when they had a 3-PT lead in the final possession. Using Markov chains to model basketball possessions is also a technique that has been explored before, as demonstrated by the study conducted by Jelaska et al. (2012), which showed how the flow of a basketball possession can be modeled as a Markov chain, ultimately resulting in a matrix for calculating percentages. However, there is a lack of research combining the use of Markov chains with actual data from NCAA Men's Basketball games. There is also little research available that considers the unique

one-and-one rule, which is exclusive to NCAA Men's Basketball. The goal of this study is to analyze FT situations using Markov chains, taking into account statistics from the 2024-25 NCAA Men's Basketball season, while considering the unique rules within collegiate men's basketball.

Methodology

Markov chains provided an effective framework for modeling basketball possessions, particularly due to their ability to illustrate changing probabilities across different states. This approach has been validated in previous sports analytics research, making it a reliable method for modeling possession-based transitions in basketball (Jelaska et al., 2012). For additional content validity, three subject matter experts (SMEs) reviewed the code and accuracy of applications in this paper. All three SMEs hold degrees and experience in computer science, data analytics, and/or sport analytics. This study is quantitative, relying on numerical data and statistical methods, and is exploratory in its goal to find optimal times to start intentionally fouling.

Prior to creating the Markov chains, eight statistics were collected to calculate the probabilities of a state changing. Data points were collected by scraping play-by-play logs from over 4,000 games of the 2024-25 NCAA Men's Basketball Season, using the CBBpy Python package by Daniel Cowan. The complete code for the scraper is available at https://github.com/anirudhsengupta/fouling-in-NCAAD1MBB. Once collected, the statistics were cleaned and adjusted to account for null values and accurate percentages for each team. A majority of null values consisted of Division 2 teams, which were removed as they were not considered in this study. The following statistics, listed in Figure 2, were collected for each Division 1 team.

Statistic	**Definition**
3-PT Rate	Percentage of a team’s plays which started with a 3-Point Attempt

2-PT Rate	Percentage of a team's plays which started with a 2-Point Attempt
Turnover Rate	Percentage of a team's plays which started with a turnover/steal prior to a shot attempt
3-PT%	Percentage of a team's 3-Point Attempts which are successful
2-PT%	Percentage of a team's 2-Point Attempts which are successful
FT%	Percentage of a team's FT Attempts which are successful
OREB%	Percentage of all possible rebounds which a team secures while they are on offense
OROFT%	Percentage of all possible rebounds off of missed FTs which a team secures while they are on offense

Figure 2: Table Listing the Statistics Compiled for the Markov Chains

The percentages listed in Figure 2 demonstrate the likelihood that the game flow will move from one state to the next. Using the collected statistics, the Markov chains for the 2024-25 Florida Gators are presented in Figures 3, 4, and 5. Each rectangle in the Markov chains is considered a state, with arrows pointing towards a possible next state. The red rectangles signify an action by the team on defense. In these weighted Markov chains, arrows are labeled with percentages based on how likely the state is to come next in the sequence of steps. For example, the arrow, highlighted in yellow, pointing towards State 6 (3-PT Make) from State 2 (3-PT Attempt) indicates that if the Gators took a 3-PT attempt, there would be a 35.9% chance that the attempt would be successful.

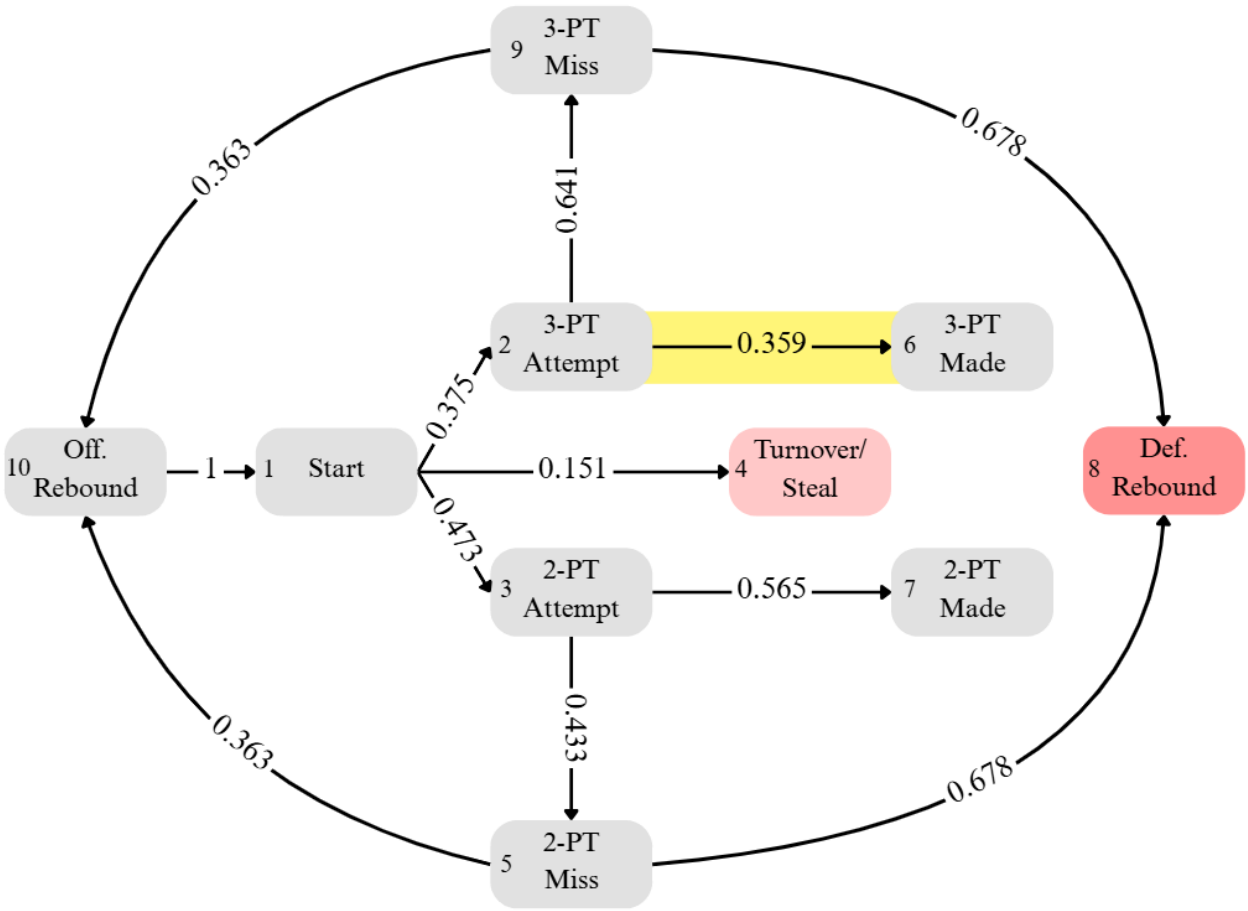

Figure 3: Markov Chain of 2024-25 Florida Gators Offensive Possession (No Foul)

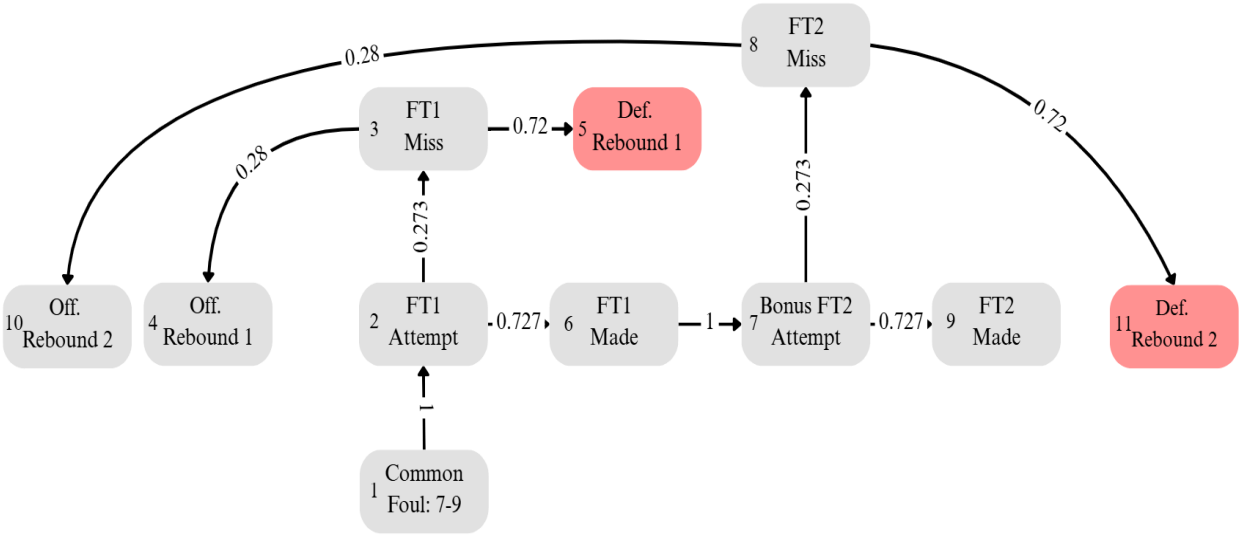

Figure 4: Markov Chain of 2024-25 Florida Gators Bonus Free Throw (One-and-One)

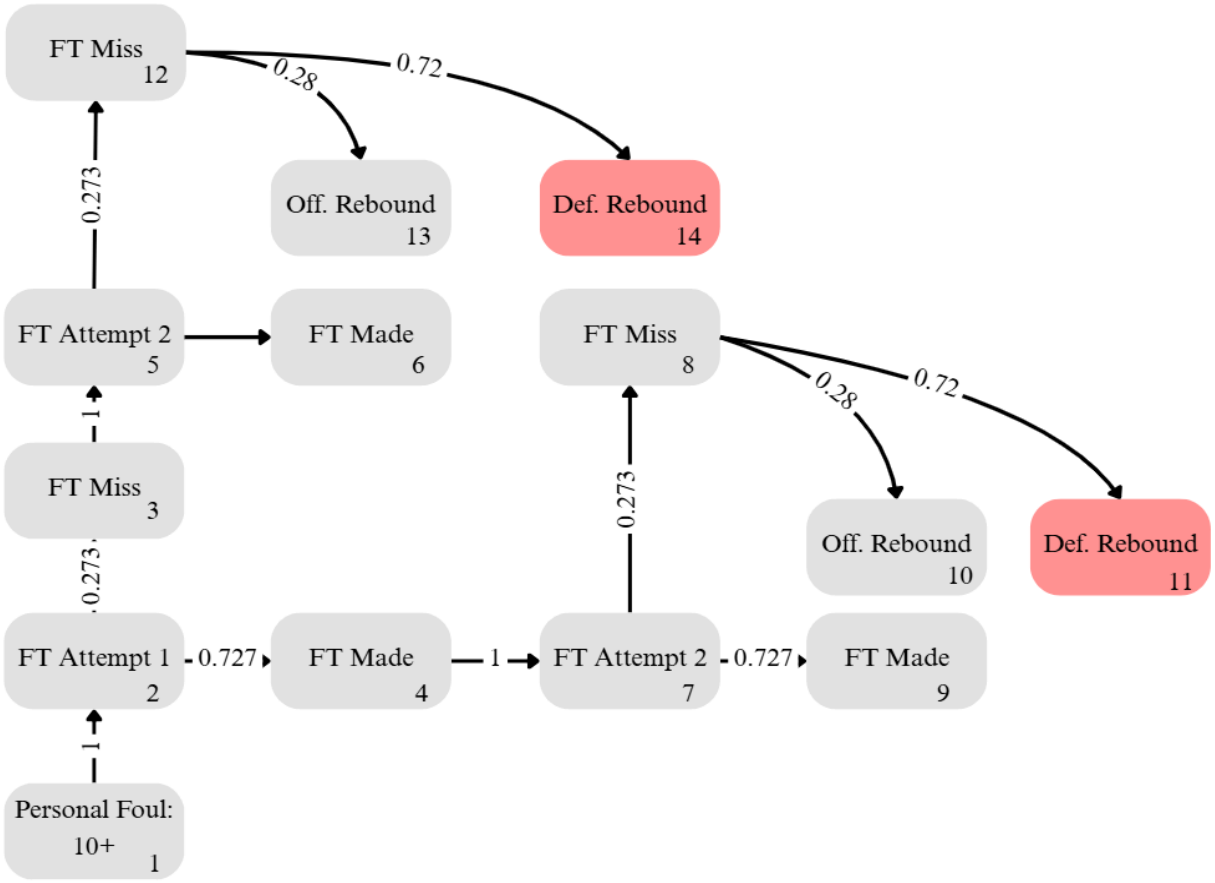

Figure 5: Markov Chain of 2024-25 Florida Gators Double Bonus Free Throw

Each chain was used to model a different scenario in a basketball game. To compare the expected points of offensive possessions and FTs, only offensive possessions that did not include a foul were considered in this analysis. For further use, the Markov chains were modeled as transition matrices, where each state has one row and one column. The probability of the possession going from one state to another was placed in their intersection, with the initial state as the row and the next state as the column. The transition matrix for the 2024-25 Gators is shown in Figure 6.

Next State

Current State

$$\begin{bmatrix} 0 & 0.375 & 0.473 & 0.151 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.359 & 0 & 0 & 0.641 & 0 \\ 0 & 0 & 0 & 0 & 0.433 & 0 & 0.565 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.678 & 0 & 0.363 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.678 & 0 & 0.363 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 6: Transition Matrix of 2024-25 Florida Gators Offensive Possession (No Foul)

To predict the likelihood of each possession ending in a specific state, the transition matrix was multiplied by a single-row matrix, where 1.0 in the first cell signifies that the possession started in State 1. Repeatedly multiplying both matrices yielded a single row matrix with the probabilities of ending possession in each of the different states. After finding the probabilities that a possession would end in a successful 3-PT, 2-PT, or FT attempt, those probabilities were multiplied by 3, 2, or 1, respectively, to calculate the expected number of points in each type of possession. To understand how the expected points from each possession changed throughout the game, play-by-play logs were scraped in five different intervals: full game, last half, last 5 minutes, last 1 minute, and last 30 seconds.

Markov chains can also be used to calculate the expected number of points when a specific player is shooting FTs. This is done by replacing the team average FT% with the player's individual FT%. Using this, I calculated each individual player’s expected points when shooting FTs and identified if a team was likely to score more with an offensive possession compared to a

specific individual shooting FTs. After the expected points were calculated for each scenario, the data were used to create line graphs to facilitate comparison.

Results

Calculating the expected points using the matrices showed interesting trends. The graphs in Figure 7 show three teams with different outcomes.

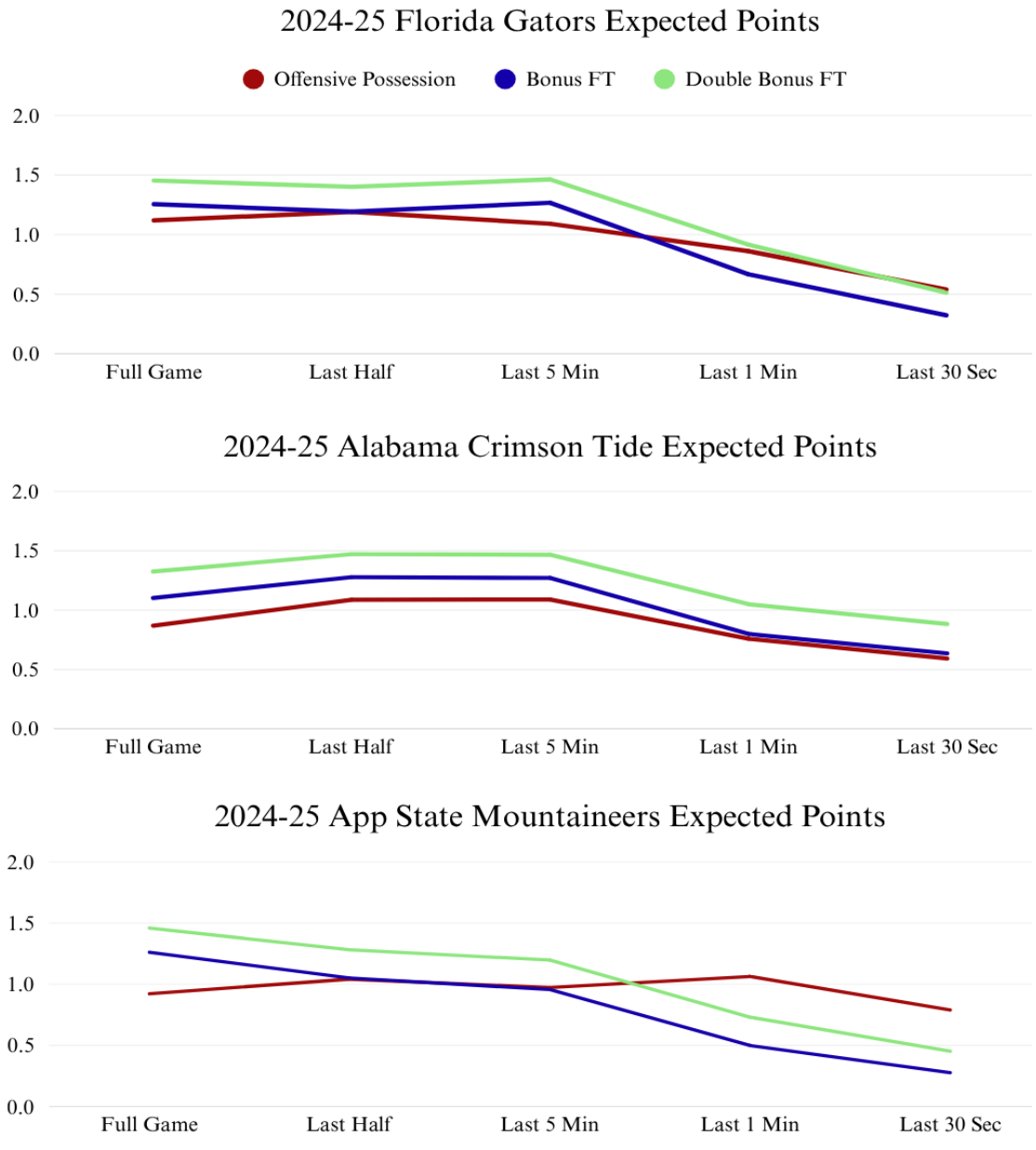

Figure 7: Expected Point Graphs for the Following Three Teams in 2024-25 (labeled): Florida, Alabama, App State

There were several common trends among the 364 Division 1 Men's Basketball Teams. Notably, for offensive possessions, there was a slight rise in expected points from the full game to the last half, followed by a steady decline as the game progressed. For FTs in the Bonus as well as the Double Bonus, the expected points tended to fall throughout the game. The intersection of the expected points for an offensive possession and a FT situation demonstrates the time at which the expected points are the same in either scenario. When the expected points for an offensive possession are higher than those of an FT situation, it is optimal to foul in order for a defense to give up fewer points.

The second application of the Markov chains was to show who to foul. Consider the Florida Gators as an example. Figure 8 shows an expected score of 1.12 points in an offensive possession when accounting for full-game statistics. By changing the value of FT% in the matrix, I found that the expected points for the Gators in a FT situation was 1.12 when the FT% was set to ~67%. This means that if the defense were to strategically foul a player on Florida at some point in the game, that player should have a FT% less than 67%.

Offensive Possession	Player 1 (87.5%)	Player 2 (84.4%)	Player 3 (76.1%)	Player 4 (60.4%)	Player 5 (58.2%)
1.12	1.64	1.56	1.34	0.96	0.92

Figure 8: Expected Points in Bonus FT Situation for 5 Players of a Possible Starting Lineup for the Florida Gators

Discussion and Limitations

This study shows that Markov chains are an effective method for calculating the expected points for a basketball team in various situations. For example, in Figure 7, the Florida Gators' expected points for an offensive possession passed that of a Bonus FT situation between the 5-minute and 1-minute mark. This demonstrates that the optimal time to start intentionally fouling the Gators would be within the last 5 minutes of a game, given that the defense is in the

Bonus FT range. The expected points from an offensive possession for the Alabama Crimson Tide, on the other hand, never crosses the expected points of either FT situation. Thus, Alabama is always expected to score more points when given FTs, making intentionally fouling a less-than-optimal strategy. In contrast, App State's expected points of an offensive possession stay relatively constant throughout the game, while the expected points off of FTs show a sharp decline after the first half. This demonstrates it may be optimal to start fouling the Mountaineers between the first half and the last 5 minutes. Observing the differences between the graphs of the three example teams shows how the optimal time to start intentionally fouling changes based on the opponent. Strategy timing insights aid coaches in effective decision making.

Along with timing, coaches look at which individual players should be fouled, as shown by the Hack-a-Shaq strategy (Young, 2021). The Markov chains highlighted specific players who were optimal targets for fouls. While the players on the court change frequently, having prepared statistics helps coaches have a comprehensive understanding of an optimal fouling strategy.

There are many opportunities to expand upon this strategy with more statistical analysis. Further sectioning of time intervals may provide even more accurate time ranges for when it is optimal for a defense to commit a foul. In the future, such a strategy could be tested through game simulations to determine its effectiveness against a team, utilizing the power of machine learning to advance understanding. In addition, this research could be expanded on to other basketball leagues and divisions, including Men's Division 2 and Division 3. NCAA Women's basketball could also be considered.

Despite the new information that the chains in this study provide, there are limitations to their use. Specifically, the chains only show expected points based on previous statistics and do not account for point differential. If a team is losing by a wide margin and time is running out,

then it may not make sense to wait until the "strategic" time to start intentionally fouling. Additionally, the expected points graphs do not account for OROFT%. For example, a team may foul the Gators and force them to shoot FTs, but Florida may get a rebound off of a missed FT and regain an offensive possession. Another factor to consider is player transfers. In college sports, athletes are allowed to transfer between schools (Johnson, 2019). As a result, many rosters can appear completely different from those of the same team from the previous year. Recently, it was announced that the Baylor Bears would not be bringing back a single player from this most recent roster for the 2025-26 season (Associated Press, 2025). In the billion-dollar industry of college basketball, research into game strategies can help coaches form effective game plans and is a field worth of further study.

References

Associated Press. (2025, June 11). *With no players left from last season's team photo, Baylor begins summer practice with new roster*. NBC Sports. https://www.nbcsports.com/mens-college-basketball/news/with-no-players-left-from-last-seasons-team-photo-baylor-begins-summer-practice-with-new-roster

Cowan, D. (2025, January 20). CBBpy (Version 2.1.2) [Python package]. PyPI. https://pypi.org/project/CBBpy/

Jelaska, I., Trninić, S., & Perica, A. (2012). Analysis of basketball game states and transition probabilities using the Markov chains. *Analysis of Basketball Game States and Transition Probabilities Using the Markov Chains*, *66*, 15–24. https://doi.org/10.5937/fizkul1201015J

Johnson, G. (2019, October 8). *What the NCAA Transfer Portal Is... and What It Isn't*. NCAA.org. https://www.ncaa.org/news/2023/2/8/media-center-what-the-ncaa-transfer-portal-is-and-what-it-isn-t

Kenter, F. (2015). *An Analysis of the Basketball Endgame: When to Foul When Trailing and Leading !* https://supermariogiacomazzo.github.io/STOR538_WEBSITE/Articles/Basketball/Basketball_Kenter.pdf

National Collegiate Athletic Association. (2024). *NCAA Men's Basketball Rules Book: 2024–25 Edition* (Version 2024–25) [PDF]. NCAA Publications. https://www.ncaapublications.com/productdownloads/BK25.pdf

Sengupta, A**.** (2025). *Fouling in NCAAD1MBB* [Python code]. GitHub. https://github.com/anirudhsengupta/fouling-in-NCAAD1MBB

Young, S. (2021, June 19). *The Science of Hack-A-Shaq: A Free Throw Conundrum*. Basketball University. https://medium.com/basketball-university/the-science-of-hack-a-shaq-a-free-throw-conundrum-855ac94565b4

Zapata-Hall, J. (2025). *College Basketball: An In-depth Study of the "Foul Up 3" Dilemma*. https://wsb.wharton.upenn.edu/wp-content/uploads/2024/12/CBBFoulUp3Dilemma_rev-uploaded-121624.pdf