TOURNAMENT PREDICITIVE INDICATORS AND TOURNAMENT SUBGAME THEORY FOR TEKKEN 7

by

Kurtis Homan

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Engineering Technology

West Lafayette, Indiana

August 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Kathryne Newton, Chair

Department of Engineering Technology

Dr. Edie Schmidt

Department of Decision Sciences

Dr. Regina Scott

Department of Technology Leadership and Innovation

Dr. Paul Griffin

Department of Industrial Engineering

Approved by:

Dr. Kathryne Newton

Dedicated to:

My Mother and Father for continued support over the years.

My Wife for keeping me grounded to all questions that arise.

My Daughters for the reminder to learn something new every day.

ACKNOWLEDGMENTS

I would like to thank my major advisor Dr. Kathy Newton for the guidance to complete my studies. My committee members, Dr. Edie Schmidt, Dr. Paul Griffin, and Dr. Regina Scott, for challenging the process and providing constructive criticism of thoughts and ideas. I also extend thanks to all the fighting game community from tournament organizers to the dedicated players that have devoted countless hours in practice to compete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

- 1B - Single
- 2B - Double
- 3B - Triple
- AB - At bats
- BB - Base on balls
- CP - Character pool
- CPC - Counter pick character
- DICE - Defense-independent component ERA
- ERA – Earned run average
- EU[W%] - Expected uniform win percentage
- EVO - Evolution fighting game tournament
- FGC - Fighting Game Community
- GRW% - Game round win percentage
- GW% - Game win percentage
- H - Hits
- HBP - Hit by pitch
- HC - Hard counter character
- HR - Home Run
- LC - Lead character
- MMR- Match making rating
- MSE - Mean squared error
- MSNE - Mixed strategy Nash equilibrium
- MU% - Matchup percentage
- OBP - On base percentage
- OH - Opposite hand
- OPS - On base plus slugging
- P1 - Player 1
- P2 - Player 2
- PCA - Principal component analysis
- Ra - Rating of player 1
- Rb - Rating of player 2
- RC - Rule of cool
- RPS - Rock-paper-scissor
- RW% - Round win percentage
- SC - Soft counter character
- SDW% - Set damage win percentage
- SF - Sacrifice flies
- SGW% - Set game win percentage
- SLG - Slugging
- SRR - Single round robin
- SRW% - Set round win percentage
- TO - Tournament organizer
- TT - Top tier
- TWP - Tournament win points
- TWT - Tekken World Tour
- UFC - Ultimate Fighting Championship

ABSTRACT

Esports have been a growing market segment for recreation and competition. Few works of research examine the decisions that competitors need to make to maximize the probability of winning. Game theory Nash equilibriums are used to evaluate options available for players to select out of game decisions related to side selection, character selection, and stage selection. Backward induction techniques are used to solve these subgame decisions. The introduction of a rating system for players is derived from traditional sport statistics. The primary factor tracked in damage dealt and damage received using the same framework from sabermetrics was used to predict outcomes of baseball games. Conclusions demonstrated tracking damage can be used to predict the outcome of a match. Other techniques such as principal component analysis did not provide adequate data to measure individual metrics for the use of predictive application.

CHAPTER 1. INTRODUCTION

Chapter 1 provides an overview of the research study that was conducted. This chapter contains elements that address the problem, identify the research question, and identify the scope and assumptions of the research. This research was conducted to analyze Tekken 7 tournament games, rating systems, and predictive analytics and their applications. This was a quantitative research study that analyzed data from Tekken World Tour tournaments from 2016 to 2020. The Tekken World Tour held 61 tournaments around the world. Of the 61 tournaments, 16 of those tournaments were held in the USA, the largest proportion of any country. Analysis of previous tournaments provided the framework to assist in the assignment of predictors for tournament placement for all players, develop a fighting game rating for each player, and perform analysis on tournament subgames. Traditional sports are used to identify how traditional sports have evolved to use statistic and predictive analytics to assist players performance and ratings.

1.1 Background of Fighting Competitions and Digital Fighting Games

One-versus-one competitions have been an attractive form of entertainment for thousands of years. Starting with training for duels and martial art sports, the idea of two individuals compete to measure each other performance of decision making or trained skills. The Romans used the colosseum to view gladiatorial fights for spectacle during their games, and new training and practical fight tactics began to evolve (Mann, 2009). As the cost of training and injuries in competition grew, sport rules began to be developed to have safe conditions for the participants. The Olympic games expanded this trend by hosting multiple countries from around the world to participate in contents in wrestling, judo, karate, taekwondo, and sumo, to evaluate the best athletes from varies background to encourage competition (Liao, 2006). More recently the

Ultimate Fighting Championship (UFC) focused on to have an anything goes fight competition from different fighting schools to observe if it was about individual player fight ability, or about a specific fighting style that had the greatest impact on the outcome of a one-versus-one matchup (History of UFC, n.d.). The UFC in 1993 used a single elimination tournament bracket that invited top practitioners of a fighting style to join. The purpose was that in a single elimination tournament would showcase fighting styles from around the world to eventually crown the stronger fighting technique. The original tournament in 1993 showcased eight fighters that used fight techniques from, savate, sumo, kickboxing, American Kenpo, Brazilian jiu-jitsu, boxing, shootfighting, and taekwondo. While some talent stood out from a physical physique standpoint, it quickly was observed that styles that utilized grappling dominated other fighting styles that focused on strikes exclusively. The technique that became the need-to-know technique became Brazilian Jiu-jitsu. The individual who used this technique was Royce Gracie and the technique that he used, that his family developed from Judo and Jujitsu sport from Japan, was the Gracie Jujitsu style (Meehan, 2020).

The early decision to select Royce Gracie to his years of dominance in mixed martial art competition was not the obvious choice compared to the rest of the Gracie family of fighting practitioners. What the Gracie family's intent was, other than win the tournament, was to show the style and technique was the most important part of the fight, while physical physique played a role, it was minor compared to the techniques used (Fusco, 2012). The Gracie family was so confident of their system, that even with a fair physical body strength, the knowledge of how to fight would minimize the Gracie weaknesses and exploit weaknesses in their opponents. A main part of the Gracie fighting style was to take the fight to the ground. By taking the fight to the ground, this limited the opponent's strengths of the 'stand up and fight' which was more popular and perceived to be the ‘way to fight.’ This perception came from that history on gladiators and sports. Boxing, Sumo, Taekwondo were all fought under rules of the sport which scored points based on strikes. The rules enforced by this individual sport eliminated the practical applications of a full contact fight, that raised the question of what if there were no rules, what fighting style would reign supreme (History of UFC, n.d.).

While the UFC is still around today, more rules have been developed to make the competition less violent from a spectator’s point of view. Even with more rules, new participants who enter the tournament begin to develop ways to either incorporate Gracie jujitsu into their own training to use against their opponent, or to develop a defense of the ‘ground and pound’ style, a style in which a fighter would bring the fight to the ground with a takedown and proceed to use strikes while in mount position. The Gracie style still is in use today, but the Gracie dominance has dwindled as more fit fighters have trained to develop their fight knowledge.

The first UFC event occurred in 1993, and 25 years later David Isaacs, the first UFC president, was interviewed for his inspiration how why the UFC was started in the first place. Part of the interview discussed the easiest way to describe the UFC was not to compare it to professional wrestling, but to compare it to Mortal Kombat the digital fighting game (Snowden, 2018). Isaacs emphasized the comparison to Mortal Kombat was to market the brutal nature of what to expect of the UFC, and not something toned down or scripted. If Mortal Kombat influences real life business and sports organization, why is there not more focus on digital fighting games tournaments.

The first digital competitive fighting game was released in 1991 titled Street Fighter 2. Upon the release of Street Fighter 2, entertainment arcades increased in popularity for patrons to have the chance to play player-versus-player in digital combat, a new technological breakthrough opposed to only play against the computer (Leone, 2014). Arcades predated home entertainment consoles in which video games could be enjoyed in their own homes, namely due to cost and technology. The digital fighting game genre was created and no longer were individuals fixed to only compete from a computer but enabled them to compete against other people. Player-versus-player creates a dynamic decision making medium that changes based on near infinite possibilities. Part of this decision making comes from character matchups and character move lists which limit options but does not give clear guidance of how options are used. The advent of fighting games created a space in which simulated combat could take place, but rather than placing emphasis on building physical strength and techniques, the physical limitations were heavily lowered, while the decision-making techniques remained. Quick decision making and fast reactions, that are required in other traditional fight sports remains a critical component on measuring the success of a digital fighting game player.

The 1990s were the birth of the fighting game genre in arcades. Starting with Street Fighter 2 in the arcades, this phenomenon helped revive the arcade industry in the United States (The Impact of Street Fighter 2, 2021). The video gaming industry had been disrupted and sales declining due to low quality and oversaturated games. With Street Fighter 2 growing in popularity, this encouraged other companies to begin creating other fighting games for consumers to play. Arcade play became a way of life for some individuals, as other traditional sports required the dedication to hone their craft. An advantage that digital fighting games had over traditional fight sport was it did not require weight classes or judges which lowered the barrier of entry for practitioners. Compared to other digital games at the time that focused on beating a high score, fighting games provided an opportunity of a confrontation of one-versus-one experience against a real-life opponent compared to a computer algorithm. As player bases increased, tournaments begun to occur to rank players. The largest fighting game tournament in current history is the Evolution Fighting Game Tournament (EVO) that began in 1996 as the Battle of the Bay.

Digital fighting game tournaments create a vast amount of data points as there are up to 100 tournaments a year spread across many different digital fighting games. The majority of all digital fighting game tournaments used the same double elimination tournament format for tournaments. Double elimination tournaments were used no matter how many or few entries there are for each digital fighting game, as it was the format ingrained in the culture. While few exceptions occurred using either round robin or single elimination, other tournament formats were rarely used compared to the double elimination standard for digital fighting game tournaments. The tournament series that set the standard for many years was the EVO, the highest prestige world fighting game tournament held once a year in Las Vegas, Nevada. The most registered participants for a single game were 3,492 entries; and the lowest registered participants for a single game was 742 entries, following the double elimination tournament rules.

Major tournaments were held throughout the year around the world while featuring a wide variety of fighting games at each tournament. A major tournament has at least 128 entries but could exceed 1,000 registered entries in a single tournament. EVO started in 2002 and moved to Las Vegas in 2005. Digital fighting game tournaments continued to expand in size which required a faster format to create brackets and advertise tournaments. Beginning in 2015, the website smash.gg began to gain popularity for tracking and sharing tournament brackets. Sharing tournament brackets, players can promote themselves and spectators observe tournament placing and seeding of players or observe their competition (Empowering esports communities, n.d.).

What is missing from the smash.gg site is a single figure to evaluate a player's rating to make accurate tournament seeding. Other issues with the smash.gg website is data accuracy. It is up the players and tournament organizers to authenticate the information is complete and reliable, however the most reliable data that is stored is which player won the match. Other crucial information such as character usage, stage usage, game record, and round record was not accurate nor available to record.

The accuracy of data appeared to be an ongoing trend that was not present in the fighting game community (FGC) which traditional sports have embraced. Baseball was a prime example of saving information from previous games. Previously recorded information allowed baseball statisticians to analyze the data and find trends. However, for many years, the findings were used in the application game play and business decisions. Baseball statistics has had a renaissance starting in 2002 when the Oakland Athletics started to use statistics to evaluate teams rather than the previously used qualitative expert opinions of old (Lewis, 2003). When the Oakland A's began to evaluate players based on their historic performance, and build a team to maximize their wins, they went on to break the amount of straight wins in a single season with the third lowest budget in the entire league. All of this was due to the use and acceptance of previously mocked statistician research beginning with Bill James, who had examined statistics to calculate and predict player and team performance (Baseball Abstract, 1981). James' research continued to be used to analyze specific situations to predict if there was an advantage or disadvantage, and baseball research continues to be a field of study for the multi-billion-dollar industry. Baseball was the proof of concept of application of theory to application and is the baseline used to research digital fighting games by identifying data to track and examining how best to interpret that data.

1.2. Purpose Statement

The purpose of this research was to evaluate digital fighting games to create a framework for data analysis and its applications. Tekken 7 was the digital fighting game of choice that is used through this research. The analysis and applications were identified as match video which aid in the evaluation of a rating system and decision theory to subgames that occur during digital fighting game tournaments. Other traditional sports have had success in the evaluation of game data that have aided aspects of the game from the players, teams, and spectators. After the identification of the evaluation of player ratings and subgame decision theory, the calculation of predictive analytics of match outcome was created as a baseline for future research evaluation.

1.3. Scope

The scope of this research was to observe Tekken 7 tournaments to establish player tournament performance metrics to calculate predictive outcome. The Tekken World Tour for Tekken 7 tournaments were majority held in the United States which has been selected for observations. Game footage from 2016 to 2020 was observed to analyze key performance indicators related to round, game outcomes, damage dealt, and damage received during game play, character selection, stage selection. These variables were then analyzed to find best fit for protection and rating of players.

1.4. Significance

The significance of the problem was to reduce the dependence on expert qualitative opinion and to shift to a focus on quantitative metrics as it relates to rating of players. Obtaining a quantifiable rating metric will reduce the time it will take to create tournament brackets using a rating system. The performance ratings that are constructed will also assist the viewership of the statistics to keep better informed on the game to retain interest. Modern approaches to matchup chart analysis are discussed, along with stage selection analysis, and player best responses for tournament play.

1.5. Research Questions

- How can a set of outcomes be predicted for a tournament first to two for Tekken 7?
- How is a performance rating calculated for individual players who play Tekken 7?
- What decisions should be made to maximize the probability of winning through tournament subgames? What are the strategies that are used for side selection and character selection with respect to the opponent?

1.6. Assumptions

- All players have acted rationally and play to win.
- Players will never throw a match intentionally or engage in match fixing.
- Players will make decisions to help maximize their chances of winning in all subgames as well as games.

1.7. Limitations

- In game strategy and tactics are not examined.
- This research only focused on decisions that can be made before a game and after games in a first to two.

- This research did not analyze player archetypes. This is the categorization of players and how they play. For example, an aggressive player or a defensive player.
- Data was sourced from public sources, so it is possible all game footage is not documented. Any footage that is not properly documented will be lost to time.
- A time constraint was added from 2016 – 2020 January because that is the timeframe starting from when the game was released to the last major tournament before COVID-19.
- Face-to-face tournaments only are analyzed because in online tournaments the rules of the game will change. To keep comparisons easier, the same rule set is used through all the game footage that was observed.

1.8. Delimitations

- Not all matches were recorded during a tournament. All game footage was searched for; however, it was up to the tournament organizers to record and store the game footage.
- This study did not focus on the individual decisions related to how to play a character, but there were a lot of decisions that had an impact on the probability of winning a match.
- Point-based systems were not under review. Each tournament organizer creates their own point-based rating system, which provides a future paper to investigate the effects of point-based leagues for tournament entry.

1.9. Definitions

- Playoffs – The top 8 players to determine the winner of the tournament

- Placement Games – games used before a tournament to decide the ordinal ranking used for tournament pairings
- Promotion Games – a game in which both players are competing for an ending position in categorical rank. Promotion implies the lower ranked player is attempting to increase the rating beyond a threshold value.
- Demotion Games – a game in which both players are competing for an ending position in categorical rank. Demotion implies the higher ranked player is at risk of losing rating points the drop below a specific threshold.
- Ordinal Rank – ranking system to assign the value of 1 to the best player and descending integers for the remaining players based on their ending position, either from points or tournament placement
- Relative Rank – ranking system that uses points or percentage to show differences in magnitude between players
- Single Elimination (Knockout) – style of tournament in which a player loses one match and is eliminated from the tournament
- Double Elimination (Knockout) – style of tournament in which a player loses two matches and is eliminated from the tournament
- Swiss System – a tournament style in which after each round a new pair is created, winners (losers) will pair with winners (losers) and this will continue for a specific amount of rounds

- Round Robin – a tournament style in which each player will be paired against all other players. Single Round Robin (SRR) is when one game is played against each player
- Hidden Rank – a computation that occurs behind the scenes and generally used for match making purposes
- Rating – a numerical representation created from historic performance information. Threshold values can be used to separate rank, for example, an ELO of 1200 is an average player, while 1800 is considered very good.
- Performance – the actual wins and losses a player or character has on historical data
- Quality Win – achieving a game or set win without losing rounds
- Win/Loss% – used to evaluate the statistical performance of a player or character
- Ranked Match – games that are played within a game's online ranked mode, most fighting games use a variation of the ELO system for match making and ranking purposes
- Match Making Rating (MMR) – the rating that is used to match players of near equal skill level
- Tournament – an event in which players follow a specific set of rules for pairing and games to be played to determine a winner. Typically, placements are made by ordinal rank.
- Character – the avatar representation that is used on screen. The character will have a preset list of actions and damage output

- Matchup Chart – a matrix created to show the historic statistical expected win rate between two characters
- Game – two players competing against each other consisting of one to three rounds (Evo Rules)
- Set – a combination of two to three games between two players (Evo Rules)
- Round – a single iteration competition between two characters (Evo Rules)
- Match – two players play a series of sets (Evo Rules)
- Player Performance Rating – a measure to compute a player's overall fighting game skill
- Hard Counter – In Matchup Chart analysis, a hard counter is the highest probability of character to win versus a known character.
- Soft Counter – In Matchup Chart analysis, selecting a Character that has over a 50% chance of winning, and taking into consideration how your opponent with counter pick afterward victory.

1.10. Overview of Study

This dissertation is arranged in six chapters.

Chapter 1 provides background information about digital fight games, decision making using game theory, and the statistics for that are to be evaluated.

Chapter 2 provides a literature review of the background information related to tournament styles, character matchups, rating systems, and predictive analytics. These topics are the foundation of building the case of analyzing the situation that is occurring related to fighting game tournament ratings. The chapter summarizes how each of the three concepts are pooled together to assist building the scenarios that were be explored.

Chapter 3 provides an overview to the methodology and framework used in this research. The chapter focuses on how to measure player's tournament rating.

Chapter 4 is the analysis that was completed for this research. A how to guide for the steps used to set up the analysis related to matchup charts and decisions players may make.

Chapter 5 explores the conclusions that can be drawn from the research. The summary statistics are shown along with final decisions that are related to the strategic choices that need to take place under the assumption of a player attempting to win as the objective.

Chapter 6 provides discussion and future research that can occur. Gaps in the research are explored as well along with potential future steps to reduce these limitations in the study.

1.11. Fighting Games at EVO over the Years

The Evolution fighting game tournament (EVO) updates the fighting game lineup every year. EVO is typically held in late July or early August, and an annual announcement event that occurs about six months before the tournament provides information about the fighting games that will be featured. The selected games are generally accepted as the best fighting games for the year, and those that are not selected, typically show the downfall of the game. Table 1 contains the list of all the titles that were presented from 2002 to 2020. Table 2 further breaks down the format of each game, if the game is either a 1v1, a sequential team fighting game, or simultaneous team fighting game. Historically, over 70% of the games entered have a 1v1 format when compared to sequential team fighting games and simultaneous team fighting games.

Table 1. List of Games at Evolution Fighting Game Tournament 2002 to 2020

Year	Games	Number of Games	New Games	Number of entries
2002	Super Street Fighter II Turbo Marvel vs. Capcom 2: New Age of Heroes Capcom vs. SNK 2	3	3	
2003	Street Fighter III: 3rd Strike Marvel vs. Capcom 2: New Age of Heroes Super Street Fighter II Turbo Capcom vs. SNK 2 Tekken 4 Tekken Tag Tournament Soulcalibur 2 Virtual Fighter 4: Evolution Guilty Gear X2	9	6	
2004	Super Street Fighter II Turbo Street Fighter III: 3rd Strike Marvel vs. Capcom 2: New Age of Heroes Capcom vs. SNK 2 Virtual Fighter 4: Evolution Guilty Gear X2 Soulcalibur 2 Tekken 4 Tekken Tag Tournament	9	0	
2005	Tekken Tag Tournament Capcom vs. SNK 2 Guilty Gear X2#Reload Super Street Fighter II Turbo Marvel vs. Capcom 2: New Age of Heroes Tekken 5 Street Fighter III: 3rd Strike	7	1	
2006	Dead or Alive 4 Capcom vs. SNK 2 Guilty Gear XX Slash Teams Street Fighter 2 Marvel vs. Capcom 2: New Age of Heroes Tekken 5 Street Fighter III: 3rd Strike Mario Kart DS	8	4	
2007	Street Fighter III: 3rd Strike Capcom vs. SNK 2 Virtual Fighter 5 Super Street Fighter II Turbo Marvel vs. Capcom 2: New Age of Heroes Tekken 5: Dark Resurrection Guilty Gear XX Accent Core Teams Super Smash Bros. Melee	8	2	
2008	Street Fighter III: 3rd Strike Marvel vs. Capcom 2: New Age of Heroes Capcom vs. SNK 2 Super Street Fighter II Turbo Tekken 5: Dark Resurrection Super Smash Bros. Brawl	6	1	
2009	Guilty Gear XX: Accent Core Marvel vs. Capcom 2 Street Fighter III: 3rd Strike Super Street Fighter 2 Turbo HD Remix Street Fighter IV	5	2	
2010	Street Fighter IV Tekken 6 Melty Blood: Actress Again Tatsunoko vs. Capcom: Ultimate All-Stars Super Street Fighter 2 Turbo HD Remix Marvel vs. Capcom 2: New Age of Heroes Super Street Fighter 4 (Women's Invitational)	6	3	

Table 1 continued

2011	Super Street Fighter IV: Arcade Edition Marvel vs. Capcom 3: Fate of Two Worlds Mortal Kombat 9 BlazBlue: Continuum Shift II Tekken 6	5	3
2012	Super Street Fighter IV: Arcade Edition Ultimate Marvel vs Capcom 3 Mortal Kombat 9 Soulcalibur 5 The King of Fighters XIII Street Fighter X Tekken	6	3
2013	Ultimate Marvel vs. Capcom 3 Super Street Fighter IV: Arcade Edition Tekken Tag Tournament 2 Mortal Kombat 9 Street Fighter X Tekken The King of Fighters XIII Persona 4 Arena	7	2
2014	BlazBlue: Chrono Phantasma Ultimate Marvel vs. Capcom 3 Injustice: Gods Among Us The King of Fighters XIII Killer Instinct Ultra Street Fighter IV Super Smash Bros. Melee Tekken Tag Tournament 2	8	4
2015	Guilty Gear Xrd -Sign- Ultimate Marvel vs. Capcom 3 Super Smash Bros. for Wii U Killer Instinct Mortal Kombat X Persona 4 Arena Ultimax Ultra Street Fighter IV Super Smash Bros. Melee Tekken 7	9	4
2016	Guilty Gear Xrd -Revelator- Street Fighter V Super Smash Bros. Melee Super Smash Bros. for Wii U Pokkén Tournament Killer Instinct Ultimate Marvel vs. Capcom 3 Mortal Kombat X	8	4
2017	Guilty Gear Xrd REV2 BlazBlue: Central Fiction Super Smash Bros. for Wii U Super Smash Bros. Melee Injustice 2 Street Fighter V Tekken 7 The King of Fighters XIV Ultimate Marvel vs. Capcom 3	9	3
2018	Street Fighter V: Arcade Edition Tekken 7 Super Smash Bros. for Wii U Super Smash Bros. Melee BlazBlue: Cross Tag Battle Guilty Gear Xrd REV2 Injustice 2 Dragon Ball FighterZ	8	2

Table 1 continued

2019	Street Fighter V: Arcade Edition Tekken 7 Super Smash Bros. Ultimate Mortal Kombat 11 Soulcalibur VI Under Night In-Birth Exe: Late[st] Dragon Ball FighterZ BlazBlue Cross Tag Battle Samurai Shodown (2019)	9	4	
2020	Street Fighter V: Champion Edition Tekken 7 Super Smash Bros. Ultimate GranBlue Fantasy Versus Marvel vs. Capcom 2 Under Night In-Birth [cl-r] Dragon Ball FighterZ Soulcalibur VI Samurai Shodown (2019)	9	1	Canceled due to COVID-19

Table 2. Categorical Count of Types of Games in Evolution Fighting Game Championship 2002 to 2020

Year	1v1	Sequential	Simultaneous	2v2	3v3
2002	1	1	1	0	2
2003	6	1	2	1	2
2004	6	1	2	1	2
2005	4	1	2	1	2
2006	6	1	1	0	2
2007	6	1	1	0	2
2008	4	1	1	0	2
2009	4	0	1	0	1
2010	4	0	2	1	1
2011	4	0	1	1	0
2012	3	1	2	2	1
2013	3	1	3	3	1
2014	5	1	2	2	1
2015	8	0	1	1	0
2016	7	0	1	1	0
2017	7	1	1	1	1
2018	6	0	2	1	1
2019	7	0	2	1	1
2020	7	0	2	0	2

1.11.1. One vs. One Fighting Game

One vs. one fighting games consist of a cast of characters to select and the characters are to face off against one another. The winner of a game is determined when one of the character’s health is reduced to zero, or when time runs out the character with the most health wins. An exception to this is in the Super Smash Bros. franchise in which each character has several stocks, or lives, and the player who loses all their lives loses, or when time runs out, the player with the most remaining stocks wins the game. Some games, like Street Fighter III: Third Strike, has additional mechanics that allow to select a special move before starting the match, this special move is selected after the opponent's character is observed. Below is the full list in alpha order.

- BlazBlue: Central Fiction
- BlazBlue: Chrono Phantasma
- BlazBlue: Continuum Shift II
- Dead or Alive 4
- GranBlue Fantasy Versus
- Guilty Gear X2
- Guilty Gear X2#Reload
- Guilty Gear Xrd REV2
- Guilty Gear Xrd -Revelator-
- Guilty Gear Xrd -Sign-
- Guilty Gear XX Slash (Teams)
- Guilty Gear XX: Accent Core
- Hyper Street Fighter 2
- Injustice 2
- Injustice: Gods Among Us
- Killer Instinct
- Mario Kart DS
- Melty Blood: Actress Again
- Mortal Kombat 11
- Mortal Kombat 9
- Mortal Kombat X
- Mortal Kombat XL
- Persona 4 Arena
- Persona 4 Arena Ultimax
- Pokken Tournament
- Samurai Shodown (2019)
- Soulcalibur 2
- Soulcalibur 5
- Soulcalibur VI
- Street Fighter III: 3rd Strike

- Street Fighter IV
- Street Fighter V
- Street Fighter V: Arcade Edition
- Street Fighter V: Champion Edition
- Super Smash Bros. Brawl
- Super Smash Bros. for Wii U
- Super Smash Bros. Melee
- Super Smash Bros. Ultimate
- Super Street Fighter II Turbo
- Super Street Fighter II Turbo HD Remix
- Super Street Fighter IV: Arcade Edition
- Tekken 4
- Tekken 5
- Tekken 5: Dark Resurrection
- Tekken 6
- Tekken 7
- Ultra Street Fighter IV
- Under Night In-Birth Exe: [cl-r]
- Under Night In-Birth Exe: Late[st]
- Virtual Fighter 4: Evolution
- Virtual Fighter 5

1.11.2. Sequential Team Fighting Game

Sequential Team fighting games are when players select two or more characters that face each other. The difference between a sequential team game to the simultaneous team game is that once a character is on the screen, they will receive no assistance from the remaining characters and will only fight after the previous the previous fight is finished. Some sequential fighting games will have game mechanics that will benefit remaining character that were selected part of the team, for instance to restore some health of the character that won the previous round, and/or a special move meter that will continue to the next character. Below is the full list of sequential team fighting games in alpha order.

- Capcom vs. SNK 2
- The King of Fighters XIII

- The King of Fighters XIV

1.11.3. Simultaneous Team Fighting Game

Simultaneous Team fighting games involve each player selecting two or more characters that are faced off each other. The simultaneous teams will have one character out on the screen as the primary character but can swap out to perform an assist move to help the character that is being used, or they can be swapped as the primary character and the bench player will be able to restore health the longer, they are in reserves up to red live which is a portion of the damage received. All the characters will share a mechanic like a super meter bar. Below is the full list of simultaneous team fighting games in alpha order.

- BlazBlue: Cross Tag Battle
- Dragonball FighterZ
- Marvel vs. Capcom 2 New Age of Heroes
- Marvel vs. Capcom 3: Fate of Two Worlds
- Street Fighter X Tekken
- Tatsunoko vs. Capcom: Ultimate All-Stars
- Tekken Tag Tournament
- Tekken Tag Tournament 2
- Ultimate Marvel vs. Capcom 3

1.12. What is a Digital Fighting Game?

What attributes categorize a digital fighting game? Digital fighting games are played between two players, or two teams. During a game, the digital fighting game is played simultaneously using a controller, with an arcade stick to control the character on the screen being the most common. An arcade controller has a lever to control the eight directional inputs that are used for movement and to perform special actions and has eight face buttons that are assigned to attack actions. Tekken 7 uses a minimum of four buttons one for each of: left punch, right punch, left kick, and right kick. The other remaining buttons can be used to map two or more buttons at the same time if the player choices too. Each player has a selection of characters, one to three characters depending on the game, to use a specific set of moves. A game is then played over several rounds until one of the players has attained the winning condition, through elimination of their opponent.

1.13. Digital Fighting Game Tournament Extensive Form

Fighting game tournaments are played in a standardized form of rules independent of the game. While there are some adjustments depending on the nature of the game, they each follow established rules that involve subgames before a game begins. The tournament game theory extensive form is as follows.

- Subgame: Side Selection
- Subgame: Character Selection I
- Game 1
- Subgame: Character Selection II
- Game 2
- Possible Subgame: Character Selection III
- Possible Game 3

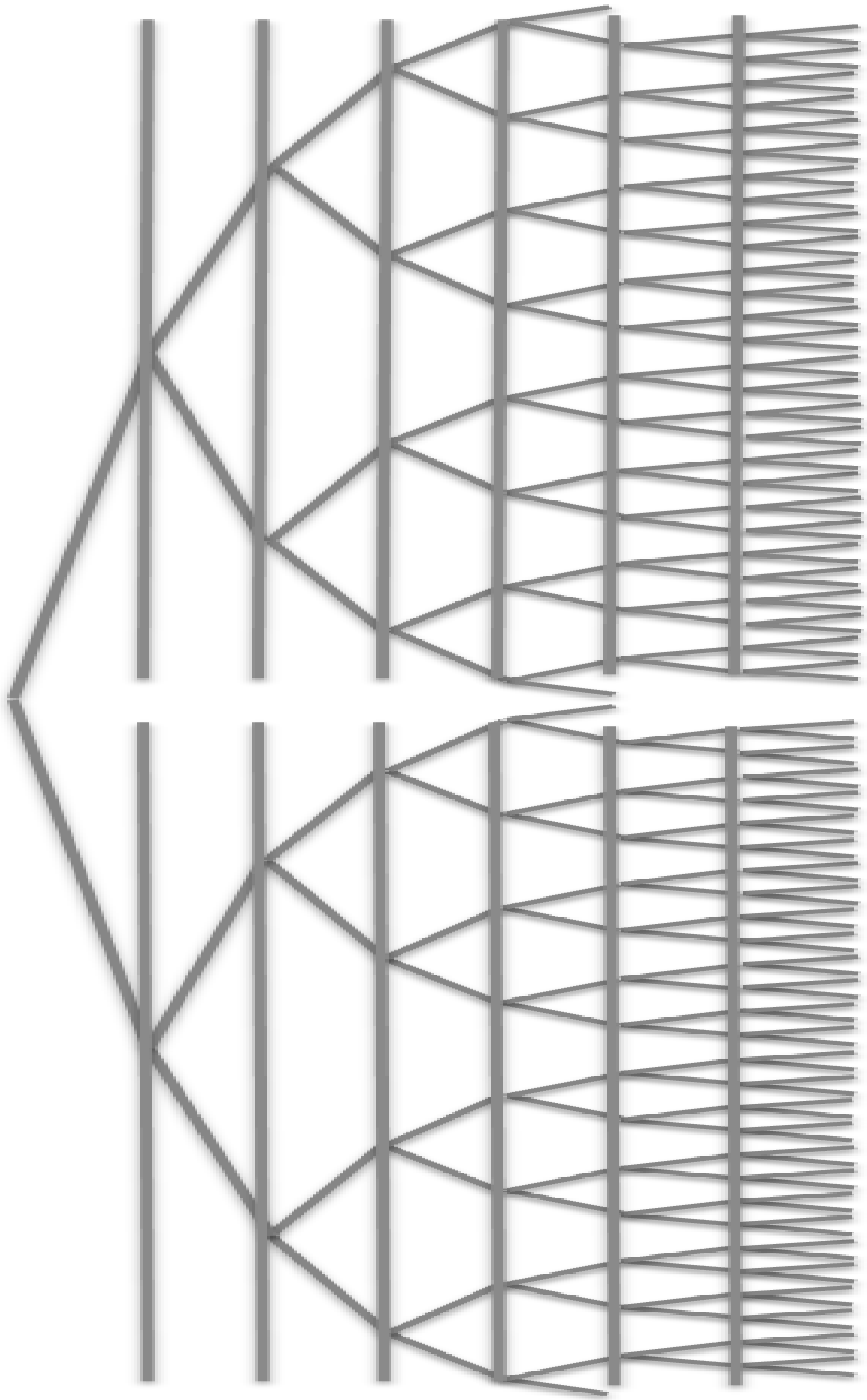

Figure 1. First to Two Extensive Form Game of a Fighting Game

1.14. Fighting Game Ladder Extensive Form

A fighting game ladder, or online play, has fewer subgames. One difference is for side selection each player can select their preferred side oppose to in person it is possible the player has to play on their least preferred side. The next subgame is typically not allowed, meaning after a character is selected the character is unable to be changed until the option to not continue to play is selected. This change modifies how you should select a character based on expected values. Finally, some games allow the player to play again, or have a dedicated stop. To clarify, some games allow players to continue to rematch in an infinite play loop, while other games have the option to continue until a specified number of games are won.

- Subgame: Character Select
- Game
- Continue to play? Yes/No (Repeated)

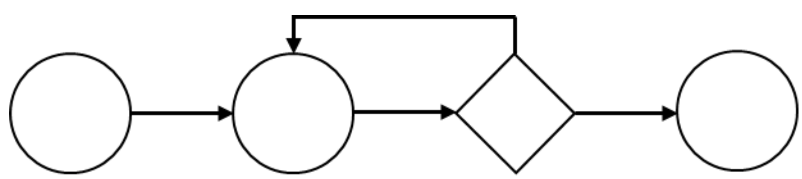

Figure 2. Online Extensive Form Game

1.15. Matchup Charts

A unique aspect of fighting games is the matchup chart creation. Tekken 7 provides the option to select from 48 characters as of January 2020. Each character has a set of abilities that a player uses to gain victory over their opponent who selects the same or a different character that has their own set of abilities. Each of these characters has a specific strength and weakness, and it occurs that some characters do not have a tool to deal with the strength that an opposing character has. This is where evaluation of games take place to measure the character’s percentage chance of winning. Fighting game players are aware character advantages and disadvantages and use the practice of counter picking to increase their probability of winning. This involves a player who has selected a character, then the opposing player then selects a character that has a better chance of winning versus the first selected character. Comparing this logic to baseball, a right-handed pitcher versus a right-handed batter typically favors the pitcher, so a defensive substitution takes place to decrease the odds of the batter making a hit.

Traditionally, matchup charts have been created exclusively by expert opinion. To date, there is no published systematic way of calculating matchup charts and finding an accurate matchup chart that uses statistical validation of all possible matchups.

1.15.1. Character Archetypes

A common way to discuss characters is to categorize on similar character fighting style or game plans. This study did not focus on categorizing these types of characteristics and is more focused on analyzing statistical matchups.

Most fighting games provide a small list of character archetypes (Ketonen, 2016). Each have specific characteristics that make them unique to the rest of the characters that can be selected. Three of the primary character archetypes are the zoner, grappler, and rush down. Games like Under Night In-Birth have further subcategories of these major categories, for instance trapper is a form of zoner. The three main character archetypes usually set up a rock paper scissor situation. Zoner beats grappler, grappler beats rush down, rush down beats zoner. A zoner typically have moves that are built to keep their opponent away, or to run away themselves, creating a cat and mouse chase. Grapplers have slower movement but once they get next to the opponent, the grappler can establish high damaging subgames loops that can quickly resolve a game. Rush downs favor speed and ways to get in close to the opponent and utilizes attacks that allow the rush down character to keep the aggression even when blocked.

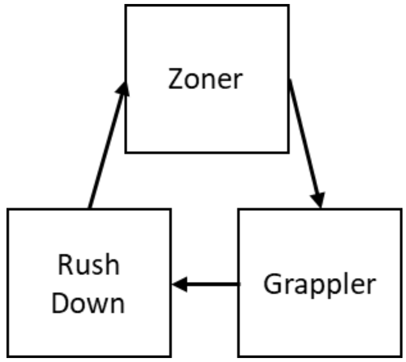

Figure 3. Fighting Game Character Archetypes

1.15.2. Mixed Strategy

After a matchup chart is created, mixed strategy equilibriums can be created to identify what is considered the best characters. This can be done by analyzing the created matchup chart. Computing the mixed strategy equilibrium provides information as to what character is considered the best, and what second or third character could be in the pool. The analysis can also identify what characters have the higher probability to see in tournaments for other players to know what characters to practice against. The matchup chart knowledge can assist with new players deciding which characters to select. This is in opposition to selecting a character for a subjective reason, colloquially known as rule of cool. A matchup chart can identify characters that have the least number of bad matchups to aid new players to focus on learning the game mechanics and work on skill, rather than having too many bad matchups and having more losses due to a unfavorable matchup. Table 3 shows a test set of data to outline mixed strategies of a matchup chart. MU% is a matchup percentage for the combination of pairings. For instance, X-Y has a calculated value of 40% probably for player 2 to win. The information can be summed using summative scale scoring, which is currently used, and the expected uniform win %

(EU[W%]) is the uniform probability of all the character equally for each specific individual character selection.

Table 3. Example: Matchup Chart

MU%	X	Y	Z	A	B	C
X	0.5	0.4	0.6	0.61	0.51	0.44
Y	0.6	0.5	0.44	0.64	0.39	0.54
Z	0.4	0.56	0.5	0.37	0.44	0.42
A	0.39	0.36	0.63	0.5	0.51	0.61
B	0.49	0.61	0.56	0.49	0.5	0.63
C	0.56	0.46	0.58	0.39	0.37	0.5
SUM	2.94	2.89	3.31	3	2.72	3.14
EU[W%]	49.0%	48.2%	55.2%	50.0%	45.3%	52.3%
AVG	0.49	0.48	0.55	0.5	0.45	0.52
MAX	0.6	0.61	0.63	0.64	0.51	0.63
MIN	0.39	0.36	0.44	0.37	0.37	0.42
BEST	Y	B	A	Y	X	B
WORST	A	A	Y	Z	C	Z
Good (60%)	1	1	2	2	0	2
Neutral	3	3	4	2	4	4
Bad (40%)	2	2	0	2	2	0

The test data can be interrupted to identify that the top characters are Z, C, and A. The insight gained from matchup chart analysis is to calculate potential character counter picks during a tournament. There are at least one counter pick subgames that can occur, with a max of three. In game theory terms, this is defined as a sequential subgame with second mover advantage. How to identify a counter pick is to identify a match up that is above 50% for a favorable outcome, and the max in a column is defined as the hard counter. A hard counter example would be if the first player selected character C, then hard counter would be to select Z to have a favorable character matchup of 58%.

Learning all the characters to play them at the highest skill is impractical as the time that would be required to practice using all characters would be infeasible in most situation. That is another use for the matchup charts, while a player cannot learn all the characters, they can learn a subset, to increase their character matchup expected value. Take for instance a comparison of a character specialist, in other words someone who only plays one character. Table 3 includes a summary of what occurs if a player that selects character Y has an expected value of 2.89. But take the same player, and have them learn a character for counter pick purposes, they pick up Z. The Z counter pick is used only for when the opponent selects X, A, or C as from Table 3 the column for Z has X, A and C to be above 50%-win rate. In other words, these matchups are favorable for Z. This increases the expected value from 2.89 to 3.48. The same logic can be used to evaluate if a third character should be picked up. Thus, the two extremes are to specialize a character, with the lowest being a 2.72, or to always hard counter, by knowing all the characters, to have a value of 3.65. In this case example, only four characters are needed to be learned to reach this highest expected matchup value, Y, Z, A, and C. Alternative counter pick strategies are discussed next.

1.15.3. Counter Pick Chains

After creating the matchup chart the next step is to identify counter pick chains. The first step is to select the first character in the list and identify the best responses for each matchup. An example from Table 3, if player 1 selects X, then player 2 will play the counter pick subgame and select a character that provides the best benefit against X. The previous matchup chart would show that player 2 will select A with a 61% chance of winning with the character matchup. To continue the chain, the next pick will provide the highest chance player will win, thus engaging into the counter pick subgame again. Player 2 would still be locked into the selection of A, which means the best response for player 1 is to select character Z. Player 2 will then select character Y, player 1 will select A, and because it is known what follows A, the counter pick chain is A -> Z -> Y -> A.

Counter pick chains are not restricted to only three characters, and it is possible to have more than one counter pick chain for each matchup chart created. There will always be at least one counter pick chain created.

1.15.3.1 Hard Counter vs. Soft Counter Chains

A distinction that should be made is the difference between hard counters and soft counters. Hard counters were described in the previous section by making a single decision to maximize the counter picking player's probability of winning by selecting a character that maximizes the expected win rate of the next match. The hard counter character selection does not consider the next match as is the hard counter should only be used during game 3. This is where soft counters come into play. Soft counters are used to look for a positive match up, like the hard counter. However, the analysis takes another step further to the potential responses the opponent will make to the initial counter. In other words, soft countering reduces the amount of risk of a counter decision, while still providing an advantage to the opponent's selection, and potential future selections.

During a game 3 scenario, the hard counter is a dominate strategy as the opponent will not have the option of counter picking your counter. How this transitions to fighting games is during tournament play during character selection III, a hard counter should always be selected to maximize the chances of winning. During subgame character selection II, there may exist a soft counter option that is preferable that will reduce the risk of the opponent selecting the counter to your counter. This risk is proposed to be solved with brute force algorithm to analyze selections of initial characters, potential hard counter chains, and soft counter chains. This information will provide more information for the players to learn and practice specific character matchups that have a higher probability of seeing in tournament. These hard counter and soft counter chains will help identify further the selection process for players that look to maximize their personal global expected win rate.

1.15.4. Multiple Character Practical Response

It is impractical for each player to learn how to learn every character in the game. That proposes the question of what is the right amount of character to learn for tournament use? Using the matchup charts to know what characters' advantages and disadvantages are helps with hard counter, or best responses, but knowing the expected mix of players and what pair of characters will net the highest expected value will provide the opportunity for at least one chance to play the counter pick subgame.

To explore this, first set up the two options for the first counter pick subgame before game 1. To begin the analysis, start with the best uniform distribution best response for counter picks. This creates the highest amount of character matchup percentage that is attainable. From there, start with a single character's expected win percentage with uniform distribution. Next look at the highest combined expected win percentage under uniform distribution by adding a character to the practice responses. This continues until all expected values are satisfied to perform perfect counter picks for each of the available character options. Below is the example uniform distribution mix with character selection in which four characters are needed to satisfy perfect counter picks for all character options.

Table 4. Example: Character Pool Subgame

CP sg	1 CP sg	2 CP sg	3 CP sg	4 PCP sg
EU[Win%]	55.2%	58.7%	60.5%	60.8%
Character	Z	Z	Z	Z
		A	A	Y
			Y	A
				C

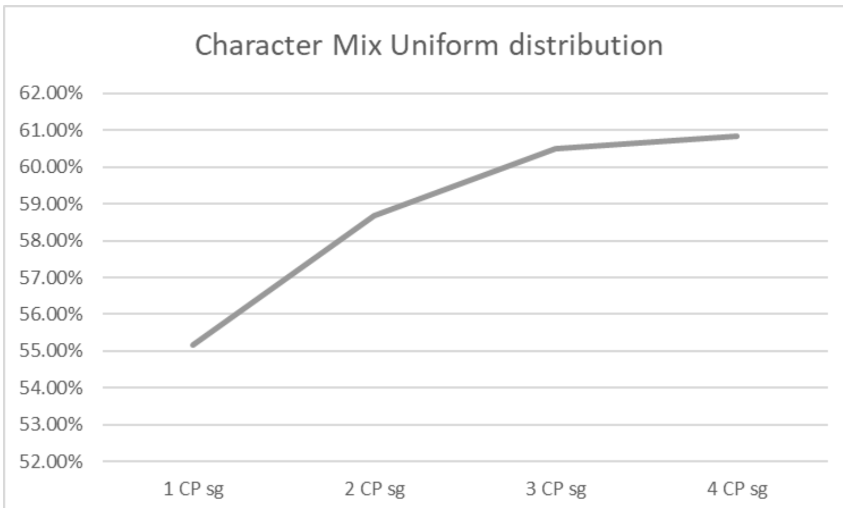

Figure 4. Graph of Character Pool Example

1.15.5. Issues with Uniform Distribution

Table 3 provided an example of uniform distribution for character matchups. The summative scale has been a standard for many years starting around the mid-1990s. However, this is entirely illogical to believe the uniform distribution is an accurate representation of the fighting game character population. For instance, in the previous section the logical decision is that a player will always attempt to maximize their percentage of winning and will not voluntarily select a character that will decrease their chances of winning. With this axiom, it is required for further analysis of the population equilibrium of the character selection to understand what an appropriate combination to practice with, and how to practice in the future.

By not taking that analysis into consideration and having the axiom of all characters will appear under the same probability begins to breakdown due to character strengths in the matchup chart.

1.15.6. Blind Pick Option Character Matchup Selection

The previous section detailed how to use the matchup chart for perfect counter picks with knowledge of what our opponent has already selected. Next is the analysis of if the opponent character is unknown, how should a player select their character? To answer this question is about selecting a lead character. A lead character will have the highest expected win percentage after the mixed strategy of character selection is calculated. Intuitively, the character selected will have the fewest number of bad matchups from the created matchup chart. If information of the character population is known, constraints can be instituted to consider smaller samples of character distribution in local scenes. This has the potential problem for players that play in smaller regions, as their matchup charts have a higher chance of locking into a local optimal mixed strategy but may not be hold true for larger tournaments.

1.16. Player Rank in Game

Each fighting game has their own internal way of ranking players. This is true for both the games themselves and for the tournaments themselves. Fighting game companies must think about the match making process to have fair matches and player motivation to continue to play. For companies this becomes difficult to access a best response, since players' perception plays a big role and consumer tastes can change. The game itself also has many different player archetypes that need to be considered, in other words not everyone is going to play rank with the anticipation of winning. Some players play for fun, self-expression, or because their friends play. These are different mindsets when compared to the player that is out to win. The situation arises in which game companies are looking at designing and modifying their games mechanics, including ranking system, to promote sales, the companies overarching goal. There is a wide variety of ways to rank, but a common method that is modified is the ELO system that was designed for chess ratings (ELO, 1978).

1.17. Tekken World Tour

Tekken 7 was released in Arcades in Japan and South Korea in 2015 with a later console release into the United States in 2016. Upon release, Namco, the designer company of the game, in 2015 supported the Tekken World Tour (TWT) to market the game. Tournaments were held with early arcade test builds of the game for players to gain interest in the game and to begin to practice their skills. When Tekken 7 became available in the USA on consoles, tournaments were still supported by Namco for marketing purposes. Starting in 2018, the Tekken World Tour greatly expanded to not only include USA tournaments, but also to include international tournaments. Namco’s official website stores information starting from the 2018 TWT where all the events were held. At the end of each year, there would be a tournament that players could only participate in if the players were award TWT points. While all the tournaments leading up to the grand finals were open double elimination tournaments, TWT were awarded for earning placement in TWT sponsored events, or for TWT Dojo events that were approved by Namco.

During the 2018 season, and all seasons prior, the only rule that was different from tournament play was that stage selection needed to be randomized. Starting in 2019, only the first game the stage needed to be randomized and after the first game, it is on the decision of the loser of game 1 to decide to either change character and randomize the stage, or to select the same character and have the option to select the stage. These rules sets have remained unchanged for the 2020 & 2021 season, other than a shift to online only tournaments after pandemic concerns arose.

The TWT provides a unique opportunity for players to not only participate in a single tournament, but to encourage players to travel and continue playing to earn season points to play in the finals. The TWT point system will not be analyzed in this research paper.

1.18. Summary

Chapter 1 introduced the research question that is addressed throughout the paper and identifies the scope, significance of the research. A complete background of digital fighting games has also been covered for pertinent material for this research. The assumptions, limitations, and delimitations are listed are referenced in later sections in this report. Next the relevant literature review will be covered to build the case for the question and the research methodology that is used.

CHAPTER 2. REVIEW OF RELEVENT LITERATURE

Tournaments and games have been a researched for 100s of years, with a large breakthrough of using economic game theory for their evaluation in the 1940s (Von Neumann & Morgenstern, 1947). Von Neumann and Morgenstern are credited with writing the text that created the field of game theory and its application. The literature review that follows holds to application of making decisions based on opponents to derive at the best response decisions.

2.1 Approach to this Review

The approach to this literature review is broken into four different sections. The first section covers digital fighting game research. Previous digital fighting game research has focused on the qualitative effects of behavior analysis and game published trade information The second section covers relevant works related to tournament formats. Differing tournament systems can impact the number of games that will take place and have an impact of time needed. The third section is an overview of game theory. Game theory is an evaluation method to explain games based on each players' best responses to one another. The game theory topics that are emphasized are Nash equilibriums and extensive form decision making using backward induction. The fourth section is about rating systems. This section will cover traditional sport games that have overlapping similarities to Tekken 7.

2.2. Previous Fighting Game Research

Fighting games have had minimum research in academia and many articles written in the community. Academic papers have focused on the qualitative nature of communities. Harper (2010) focused on evolution fighting game tournament and the community surrounding it.

Harper continued his research of the toxic behavior that exist in the community. Themes of gatekeeping, bullying, and other behavior that keeps players away was a focus. Harper separated categories into the top competitors that will do anything to win, and those casual fans. Harper (2013) continued researching this topic in 2016 with an update of the community.

2.3. Types of Tournaments

There are many different types of tournaments that exist to run and determine a winner. Some tournament styles are discrete in how they are set up, while others are adaptive as games are played. When it comes to tournament selection, there are many elements that should be taken into consideration. Variables to consider are number of registered entries, total time available, pairing and seeding, space of venue, number of play stations, cost of rental, price of tickets, energy requirements, internet requirements, and other equipment need to run a tournament. For a selection process, the variables that matter the most are registered entries, total time required.

Knowing total number of registered entries, or a maximum amount, is the first variable to understand to hold a tournament. The number of registered entities has a direct impact on time and other factors to consider. The next variable is time allotted for the venue is what it sounds like; the cost to rent out space for an extended period is costly for TOs. Knowing the time constraint helps establish what tournament format should be used as there is a tradeoff of the amount of pairing comparisons and total time. For instance, a round robin tournament may be more accurate, but the time allotted for each of the participants to play one another would break the time constraint allowed.

Single elimination tournaments are tournaments that are arranged in a way in which each participant plays against another until they suffer one loss, then they are eliminate from the tournament. March Madness, a basketball tournament for colleges in the USA, is the one of the most known form of a single elimination tournament. The seeding for these tournaments is arranged in a way that during the first round of tournament play matches up the highest ranked seed versus the lowest rank seed. The opposite side of the tournament bracket will have the second highest rank seed versus the second lowest rank seed. This seeding practice follows this system. The idea of the single elimination bracket is to have 32 teams, and the first four should have the rank 1 versus rank 4 and rank 2 versus rank 3 to ultimately have rank 1 versus rank 2. Criticism of the tournament is that this is a king-making tournament from the player's perspective. Particularly if during the season or previous tournament the highest rank teams have a distinct advantage over their first few rounds in the tournament, while weaker opponents are at a distinct disadvantage. Players that are caught in the middle would then show they have the fairest matchups early in the tournament. In a repeated style of tournament, this style of tournament would replicate the nearly same top teams consistently making the tournament finals again and again unless something went horribly wrong.

Double elimination tournaments are the current standard style for fighting game tournaments. The initial set up is identical to the single elimination tournament, but it takes two loses to be eliminated. After a player loses their first game, they are placed in the losers' bracket which functions as a single elimination. The difference is there is a lot of uncertainty about possible future opponents and there is a lot of down time waiting for players to play through the winner's bracket, to then form the losers' bracket which ultimately have the winner of the loser's bracket face off the winner of the winner's bracket. When this occurs, the winner's bracket has an advantage since they will only have to win one game while the final loser must win two games to be crowned the winner. Other Olympics sports, like judo, have a variation of the double elimination, except when you enter the losers' bracket, you cannot be deemed the champion (consolidation bracket).

The double elimination tournament reduces risk or variation for favored players while an unfavored player will have a harder time to win two unfavorable matchups compared to one.

Round robin tournaments were created to have everyone be paired off with one another. Round robin wheels are created to create pairing that work with even and odd number of players. A big negative with round robin tournaments is the amount of time that is needed to complete the entire tournament. Another issue with round robin play is what to do with tie breakers. Traditionally, tie breakers in a round robin will use statistics and rate quality wins. Quality win tie breaking is very controversial, and alternative methods generally need to be accepted by the game players in general. Needing alternative rules for this can be confusing, or it can be a method to increase tension in a mini playoff, however this extends the number of games played to an already long format.

2.3.1. Four Main Tournament Styles

Tournaments can follow many different rule sets when it comes to the style or format. There are four main styles for tournaments that can be used. The single elimination, double elimination, round robin, and Swiss tournaments. Factors such as time per match, time for the venue, and player fatigue play roles intro selecting a tournament style.

Single elimination is the style of tournament in which each player plays until they suffer one loss. As soon as a player loss occurs, that player is eliminated. This type of tournament is arguably the easiest to set up. It is not as time intensive to run the tournament, or the manage the operations. Based on the nature of losing one game and eliminated, this tournament has the largest variation when it concerns player performance.

Double elimination is the style of tournament that works like a single elimination tournament, however once a player loses their first match, they are placed in a losers' bracket and play until they receive their second loss at which point you will be eliminated from the tournament.

Round robin style has each player play against one another. In this style, it is known that each player will play a specific number of matches. Even if a player loses all their matches, they will continue to participate in the tournament until the final score. The winner of a round robin is typically the player that has the most wins, but it is possible that more than one player is tied with the highest number of wins. At this point, the tournament should have additional rules for breaking ties. The two common ways to resolve this is one, have the tied parties play one more game winner takes all, or two is to use arithmetic of the matchups to determine a winner. The second option is to observe all the matchups and score a separate point value and for each loss subtract a value, this ending value is then used to determine the winner. If the game has alternative scoring in the game, a rule could be written to take total points scored into consideration if the win and loss is a tie.

Swiss style tournament has a wide variety of research on how the Swiss style is used (Duke, 2018; Team, 2020). A Swiss style tournament begins with an initial pairing and after the wins and loss occurs, a new pairing is made with the same running score. What this means is if you won your first match, you will be paired with an opponent that has the same or similar record. This repairing continues until the number of matches are successfully played. The number of matches that are played are described in the rules by the tournament organizer, however there is a correlation between the number of players signed up and the number of matches that will be played.

2.3.2. Three Minor Tournament Styles

Alternative methods include consolidation tournament, toilet bowl, and snake. A consolidation tournament can be like a double elimination tournament. It is a separate style of tournament that is started as soon as a player is eliminated, a consolidation is started. A consolidation tournament can operate then as a single elimination tournament; however, the winner of the consolidation tournament then does not play against the winner from the winner's side tournament like in a double elimination tournament.

A toilet bowl tournament is less common to be used in practice. How a toilet bowl works is it takes the bottom number of players and after pairings are created and the winner of the match leaves the tournament, the loser is paired up. A toilet bowl tournament operates like a reverse single elimination tournament. At the end of the tournament, there will be one single player that did not win a single match in this reverse single elimination tournament.

The last alternative method that will be discussed is the snake tournament. The snake is also not overly popular. A snake tournament is a mixture of round robin and Swiss style. A round robin tournament you would know that you play all the players, and you would know the order in which they will be played in. A Swiss style makes new pairing after each round, so you do not know who your next opponent will be, and there is a fixed number of opponents that competition will take place. The snake method takes the round robin and assigns a specific number of matches that each player will play. These pairings are created and posted ahead of time, so rather than in a Swiss style must watch until all players have played, and a new pairing is created, the snake style will be able to enable players to find their next known opponent and begin a match when both players are available, and a setup is ready to be used.

Table 5. Tournament Format Matches and Bracket Rounds Equations

Format	# of Matches	Min Bracket Rounds
Single Elimination	N-1	$2^{n-1} < N \leq 2^n$
Double Elimination	$2 * N - 1$	$2^{n-1} < N \leq 2^n$ $then$ 3*n-1
Round Robin	$\frac{N * (N-1)}{2}$	$\frac{N * (N-1)}{2}$
Swiss style	$\frac{N * M}{2}$	M
Consolidation	$2 * N - 2$	$2^{n-1} < N \leq 2^n$ $then$ 3*n-2
Toilet Bowl	N - 1	$2^{n-1} < N \leq 2^n$
Snake	$\frac{N * M}{2}$	M

2.4. Tournament Pairings

Until now only tournament bracket styles have been discussed. To fill those brackets there must be a system to be used for pairing. There are a few commonly used systems for creating pairings for tournaments, each with their own advantages and criticism. The pairing methods are the king making, McMahon, and aggressive pairings.

King-making pairing is what is used in the traditional brackets. The king-making system is used intentionally for the purpose of having early bracket rounds to be mismatched and focus on the final matches. The mismatch occurs with the pairing of the best player to the worst player, an example can be found in Table 6. The purpose of early mismatch of skill is to increase the probability of making the final matchup the theoretic most entertaining, and fair matched, game to watch of the tournament.

Multiple repeated games are the weak point for king-making. Lower ranked players have little incentive to continue to work on improvement as they will only face off versus the best opponents first, which will require the player more effort to show improvement. This comes in the point of knowing they will have an unfair match in later tournaments. The rational play in this case, is to not play the game anymore. From a longevity standpoint, this damages a game's potential longevity of this repeated game as the number of players typically decrease year after year.

Table 6. Example: King Making Pairing

Initial Match
1 V 8
3 V 6
4 V 5
2 V 7

McMahon pairing is in which the skill gap between each matchup is minimized in the beginning tournament rounds. The McMahon system will select pairings in a more uniform manner, while still providing an advantage to a single player based on skill. A difference between the king-making and McMahon matchmaking is that lower ranked players have a better chance of upset especially earlier on as the skill rating is not as drastic as when compared to the king making pairing.

A repeated McMahon tournament provides more motivation for both low and high ranked players. The earlier rounds will be easier, they are not as easy as with king making. The benefit is the motivation from the lower ranked players to come back and having a better pairing in the early rounds before the player is potentially eliminated. Rather than the lowest ranked being matched up with the highest ranked, there will be better matchups to shift slowly up or down, rather than being pummeled in the first few rounds only to return there.

Table 7. Example: McMahon Match Making

Initial Match
1 V 5
2 V 6
3 V 7
4 V 8

A controversial pairing method is the aggressive has advantages. Aggressive pairing works like McMahon, but the first pairings can have the 1st rank and 2nd rank play in the first round. This brings the controversy of the possible most interesting match of the tournament to occur in the first round. Since the most interesting match has already occurred, there is little incentive for spectators and sponsors to continue watching the rest of the matchups.

A repeated aggressive pairing game provides early rounds to be exciting, and the best scenario for unfavored players. However, favored players, especially near the top, will have greater shifts in top ranking placement. The worse offender is the longer the tournament bracket continues, the quality of matches gradually becomes worse due to higher ranked players being eliminated earlier in the tournament.

Table 8. Example: Aggressive Pairing

Initial Match
1 V 2
3 V 4
5 V 6
7 V 8

2.5. Tournament Scheduling

Academic research about tournament style have mainly focused on scheduling problems for tournaments. This comes in the form of how a round robin tournament should be scheduled (Trick, 2002; Henz, Müller, & Thiel, 2004). These papers introduced terms that are used in round robin play like SRR for single round robin and DRR for double round robin and measure the effects of which is best under different time considerations. DRR will be more accurate in ranking all the players when compared to SRR, just from having more matches recorded and confirming round 1 win/loss.

Other research analyzed single and double elimination tournaments. Deck and Kimbrough (2015) investigated these tournaments from the player perspective of the behavioral effects of engaging in elimination tournaments and making the decision to intentionally lose a match, to have an improved future matchup because of the loss. Older research took on the behavioral aspect from a mathematical approach to determine when it is a better decision to throw a match in some cases to play against easier opponents for the remainder of the tournament until the finals. Economists Koller and Pfeffer (1997) investigated behavioral aspects of why players participate in tournaments. Koller and Pfeffer identified the incentives of a player that was sponsored or when a player wanted to enter the game to for the enjoyment factor. The dichotomy of players that are there to win shows that there are another group of players that can be incentive to participate, if the tournament organizers can identify what that is for the outlying group.

Other research analyzed specific games that used tournaments. Ben-Naim et al. (2007) compiled their work on college basketball and how it grows and what changes needed to be made to organize the pairings and styles for the teams involved. NASCAR was a focus for Becker and Huselid (1992) that introduced the SPREAD idea. The SPREAD discussed that as the number of entries into the game there are, the more payoff positions that should be. The research also examined the pricing aspects for spectators and how it should scale in relation to the popularity of the sport.

Connelly et al. (2014) focused on tournament accessibility. The research “Tournament theory: Thirty years of contests and competitions” focused on tournaments in the general sense of business rather than only on games. This work provided an eye-opening framework for how to build a tournament style and how it will affect the players that will participate in that tournament. Style papers continued to look at other aspect of weighted or handicapped tournaments. Shepanik (2015) used graph theory to measure the effect of a tournament style had on a weight pair. Primarily this research was directed at identifying an initial tournament style and pairings that would evolve over time.

2.6. How to Handle BYEs, No Shows, and Early Drops

Problems that occur no matter the tournament style are how to handle situations of BYEs, no shows, and ties. BYEs occur when there is only one player to a pair. The situation occurs most commonly when there are an odd number of players. A tournament rules should determine how byes are handled, especially when wins and losses are used for tie breaker purposes.

No shows are when a player is signed up and should be there to play but is not. Most instances of no shows are handled as a bye; however, some tournaments have additional penalties to players that do not show up, or even show up late.

Individual played games, a common phenomenon that occurs are BYEs, no shows, and early drops. BYEs occur when there is a match that is scheduled, but the opponent does not show up, so an automatic win is rewarded. However, in some tournament systems, this can hurt a player that receives the BYE in the current tournament, as well as future tournaments that use the same rule set based on performance.

With the rise of video games that are used in tournaments, there have been many different variants of how to run the tournament. While double elimination seems to be popular now, there are other methods that may be better from multiple perspectives that are lost using traditional double elimination rules. This research was conducted to explore different ways to run a tournament from a new player, professional, spectator, and commentator perspective.

Methods of running a tournament will be described to examine aspects of time the event organizers have to run the tournament, the number of stations that are available (to measure concurrent games), entry numbers, and ranked and unranked players.

2.7. Traditional Sports and Games

Majority of sports and spectator games, there is a season of play, and then a final playoff. A few examples are Major League Baseball has is the World Series, National Football league has the Super Bowl, College Basketball has March Madness. Of these three examples, College Basketball provides the most interesting overlap of tournament styles and pairings. Each of these sports have a regular season that is played through a period all up to a final match that draws the most attention. This final match in many cases is fought among the highest ranked teams with the least amount of difference between team skills. Every game has a different method of playing against one another as there are multiple different constraints that effect the nature of the game. A MLB season plays typically three games versus the same opponent with a total of 162 games played during regular season. The NFL typically plays 16 games during the season, playing against half of the football teams in the league. College basketball plays in the range of 25-35 games person season depending on the division. In each of these examples, a constraint that is used for game matchups are geographic location to create divisions. These divisions make it less costly for individual teams in terms of travel to face one another, with the occasional matchup that is played outside of the division. These divisional teams near the end of the season will then select the best performing team in each division to be selected to be played in the playoffs for the conclusion of the season to crown a champion.

The purpose of the seasonal play is to gain more information of the teams, broken down by division and/or location to partitions the games. Round robin style within the divisions, followed by a few games that will cross the division game. Ultimately, the top of the divisions is selected to be played in the playoff finals.

Depending on the game, the lengths of seasons, number of games played, and the number of teams differ. Rules differ as well when the teams are not locked to a specific country but are open to international play. Due to logistics, teams traveling can make playing more games exceedingly expensive for travel and room and board. This is the case where fewer games are played, but the teams in the international division are lower. In other words, for international divisions, there are smaller divisions that feed into the international ones. This enables the feeder divisions as an initial step to filter out players and teams to have a lower rating, to allow the higher rated player to compete with another.

Seasons serve an important purpose, but when an open tournament is assigned, conditions of the game need to consider what is defined as a season. With open tournaments, the World Poker Tour is another form of open tournament to assist in having other tournaments throughout the year to open to the World Series of Poker. The accumulated points that are awarded through the tour are used for World Series of Poker for acceptance into the tournament. The point system rewards players that can participate in tournaments throughout the year and providing information for the tournament organizers of previous games how they performed.

2.8. Game Theory

When economic game theory is discussed, a common principle is the Nash equilibrium. Originally published by Nash (1950) about the technique, it has influenced many researchers and practitioners to take note on how individuals will respond in non-cooperative games. The Nash equilibrium is set up to draw a matrix of possible options. In a two-player game, best practice dictates decisions for player one is in the rows, and decisions for player two will be the columns. Each cell in the decision matrix is filled with different payoffs for each individual player. The formed payoff matrix is then used to evaluate decisions to identify dominate and dominated strategies. An equilibrium is found within the system when both players would converge to a single decision pairing with respect to each other's decision.

Further study of economic game theory is Gibbons is a modern tool for an overview and learning game theory (Gibbons, 1992). All further research mentioned has some attributes to Gibbon's work. Maskin (2011) researched the theoretical and application shortcomings of Nash equilibrium, however Maskin was able to argue that the drawbacks of the logic of mapping out n-player responses to a specific game are too few when compared to the applications of applications of game theory. In short, a Nash equilibrium analyzes an n-player game and using a payoff matrix maps out all the possible outcomes from each players selection and responses. From the created payoff matrix, it is rationalized to find an equilibrium point, or points, that exist within the game the mirror the application of standard play.

An alternative to the standard Nash equilibrium is when there is not a dominate equilibrium, but a mixed equilibrium to the decisions needed to be made. A classic example is the soccer penalty kick game that needs the appropriate mix of penalty kicks to the left, or to the right. Mapping out this repeated game shows the probability that each selection should be made and is called the mixed strategy Nash equilibrium (MSNE). While mixed strategies have been in use, the application has been tested included as recent as Azar and Bar-Eli (2011) analyzed a sample size of soccer penalty shots using the MSNE and other probabilistic models. It was concluded that the MSNE was a better predictor compared to other prediction methods.

The next natural progression for Nash equilibrium was previous games assumed to have complete knowledge of all the players' actions and their payoff matrix to make informed decisions. Further research examined how Nash equilibrium analysis changed when all payouts were unknown based on the individual players utility of the combined known available actions. This incomplete information comes from the private value of an individual player on their own specific payoff of a specific action. Rabinovich et al. (2013) used a fiction play algorithm to investigate how to implement such an algorithm to find a solution to game in which private value of payoffs were unknown, but the actions were known. Through their research they were able to find pure strategies where the primary game used in their research was a double-blind sealed auction bidding game. A double-blind sealed game involves players to make a hidden bid to when an auction. Typical bids are won with bidding the highest amount, but that bid is built on each bidder's intrinsic value of the private good they are bidding on. In the case of a house, the house utility will change based on everyone who can utilize the house if won.

Within any variation of the Nash equilibrium, that be a single equilibrium, mixed strategy, or Bayesian Nash equilibrium, a subgame can occur at a cell in the payoff matrix that needs to be solved before completing the solution space of the parent payoff matrix. The steps to solve a subgame are the same as to solve a Nash equilibrium itself. It is important to understand that the rules of the subgame are defined the same way, but after an equilibrium is found, the subgame solution is then substituted into the main game that is being analyzed. In other words, a Nash subgame will have to be solved first and back tracked into the primary game to solve that.

It is possible to have multiple subgames in each other, and the same process of solve the smallest game first to solve the next subgame etc. is used until the primary game has a completed payoff matrix.

Several subgames can occur, or multiple decision points can occur during a game. At times it is easier to be displayed in an extensive form game. An extensive form game is laid out in a decision tree that is comprised on nodes and edges. Each node is a decision point while the edges are the decisions or outcomes that can occur. Once a game is mapped out in extensive form, the next step is to use backward induction to solve the game. Backward induction involves evaluating all the terminal decision points in the extensive form game to select the series of actions that yield the best result. Aumann (1995) in his research was able to prove that “if common knowledge of rationality obtains in a game of perfect information, then the backward induction outcome is reached.” This is further proof that if public knowledge can be identified and mapped in an extensive form game, and if all players act rational, the found conclusion will be the best response.

The extensive form game also provides that benefit of evaluating first mover and second mover advantage games. Another classic game theory game is the game of nim that researchers started using mathematics to explain a solution (Bouton, 1901). The nim game involves players that take 1 to 3 sticks at a time, and the player that takes the last stick loses. While there are mathematical formulas that explain the algorithms, game theory extensive form displays the information in an easy to understand visual of the best course of action and sets up the framework of first mover advantage. The game of nim has a first mover advantage based on the sum of the sizes of heaps is not zero otherwise there is a second mover advantage. Because the size of the heaps determines which player was able to take advantage of the constraints of how the game is played to determine the player advantage for the game. If the strategy that is taken to reach the finite known number of sticks, advantage player will always win as this is not a game of chance. It is to extrapolate, if the extensive form game can visualize the series of decisions that need to take place and to identify if there is a player that has a first or second mover advantage, then the practical applications of the extensive form should be used.

2.9. Game and Character Balancing

There has been little published literature on this subject. Game developers generally only address questions about game and character balancing via FAQ sessions. However, Sirlin, founder of Sirlin games, wrote a series of articles about game design and balancing. The articles include: Definitions (2014), Viable Options (2017), Fairness (2015), Intuition (2014), Game Balance and Fantasy Strike (2019), Game Balance and Yomi (2015), and Solvability (2014). Much of the topic is how to balance the game of Yomi that is a card game that's purpose is to mimic fighting games. Sirlin's background comes from competing in tournaments and then he went on to develop Street Fighter 2: HD Remix, that was a featured game at EVO in 2009 and in 2010. Sirlin then continued to develop current game of Fantasy Strike, using the same characters from Yomi, into a simple by design fighting game. Fantasy Strike has not been a featured game at EVO but has made appearances as side tournaments and showcases for the game.

The entries consist of five papers that describe designing in a general sense, and he has further articles on individual games like Fantasy Strike and Yomi that he was the lead designer on. Sirlin grants insight on how game developers look to balance their games. There were some unfortunate flaws in the papers though, which showed a gap in knowledge of these theoretic principles (Solvability, 2014). Sirlin identified how to exploit in the game of rock-paper-scissors (RPS) in which a player is known to select Rock 100% of the time. Sirlin then stuck with the optimal strategy of selecting paper, the only option that beats rock, only 33% of the time since that would be the optimal value from the pre-calculated mixed strategy Nash equilibrium. The breakdown occurs as Nash equilibriums are formed with the assumption of knowing, or unknowing, what the opponent will do, and making the best selection. Taking that strategic step, if it is known that the opponent will select rock, it is optimal for us to select paper. The concept of conditional, or Bayesian Nash equilibrium, takes this thinking into consideration. While a global stable environment for RPS is to evenly select each option, the Bayesian Nash equilibrium considers an individual behavior to make decisions that maximizes the expected value payoff.

In the same article of "Solvability" Sirlin discusses skill and makes it a point for several paragraphs that knowledge of the game is not a skill. Sirlin references chess end games and opening of memorizing, i.e., studying the game, and that knowledge is not an inherit skill. To the same degree, any education is not a worthy skill to pursue by this way of thinking which makes the author identify this argument as a flawed argument. Knowledge about an opponent also implies a removal of skill, even though he previously was making a point that it is important to adapt to your opponent. Sirlin finished the article with the example of another game he designed that included randomness in its game mechanics and discussed that to keep a game interesting it must be difficult to solve a pure solution and goes to show that some players in his game spent pages to solve a subgame thus making a good game. The main point was that no dominate strategy exists. Mostly this can be summed up by game theory expressing right away you do not play a game with a dominate strategy, and the author believes that is what Sirlin was attempting to say with this article, rather than Sirlin's opinion about memorization and pure strategies.

Next was Sirlin's four-part series on game balance. Each article was written with a focus on a topic that included: Definitions, Viable Options, Fairness, and Intuition. Sirlin's first article was about definitions (Balancing Multiplayer Games, Part 1: Definitions, 2014). Definitions that related to this study included balance, viable options, fairness, symmetry, and asymmetry. An important take away is that Sirlin identified that fighting games by their nature provide an unfair environment. Once players have locked in their character selection, they are then locked to that specific character's strengths and weakness of options that become more or less viable compared to the opponent's character selection.

> "A multiplayer game is balanced if a reasonably large number of options available to the player are viable—especially, but not limited to, during high-level play by experts.
>
> Viable options: Lots of meaningful choices are presented to the player. They should be presented with enough context to allow the player to use strategy to make those choices.
>
> Fairness: Players of equal skill have a roughly equal chance at winning even though they might start the game with different sets of options / moves / characters / resources / etc.
>
> > Symmetric: Same starting options
> >
> > Asymmetric: Diverse starting options"

Sirlin's second article (Balancing Multiplayer Games, Part 2: Viable Options, 2017) begins with introducing his termed Yomi Layers. In essence, Yomi Layers are the mixed strategy decisions and the evaluation of these mixed strategies to determine if there is a dominate move, or if a situation has a mixed strategy equilibrium. For example, Sirlin placed emphasis on Yomi

Layer 3, which can be drawn to a 3x3 matrix, and from a design standpoint he says to not worry about Yomi Layer 4, a 4x4 matrix, as he believed player will resort back to their original best most, the move that should be used with the appropriate highest percentage. Sirlin identified if a situation exists in which a dominate strategy is identified, in other words a situation in which this is only one decision that has no counter, then game balance should be used to eliminate the dominate strategy in each situation. An important distinction Sirlin made is the difference between local and global balance. Sirlin defined local balance as a specific situation, while global balance is the result of a match. Certain local situations can be unfair, but if overall the character matchup is unfair, on a global level, then there needs to be adjustments. Another example is given that used concepts of probability. Sirlin provided the example for a character matchup between E. Honda and Guile in which Honda can be in a checkmate position, a situation that Honda should nearly win 100% of the time, but the probability of getting Guile into that position is very low. To express using Sirlin's terms, Honda had a dominate strategy locally, but globally this is fair since the chances of Guile getting in the corner vs. Honda is extremely low.

Sirlin continued in his article "Balancing Multiplayer Games, Part 2: Viable Options" to discuss decision making and using game space to create interesting play situations. Later he pulled influence from game theory by comparing the design decisions he made for his games of Kongai and Yomi from the classic game theory problem of the prisoner's dilemma. However, prisoner's dilemma is a simultaneous cooperation game, not a zero-sum competitive game. In fact, prisoner's dilemma does have a dominate strategy, something that Sirlin has discussed to avoid in game design. This means the comparison is only with a simultaneous play, and no other features of the prisoner's dilemma. A better analogy is using game theory in the game of penalty shots, in which one player must decide between kicking left or right, and the defender either must defend left or right, and this creates a mixed strategy. Another example is one Sirlin used before of rock-paper-scissor, however prisoner's dilemma is a very different game.

Sirlin ends the discussion about playtesting that comes from either data or expert opinions for finding strategies. Sirlin identifies that new players, even previous experts, my overlook global strategies since they get locked into their initial strategy local equilibrium. Sirlin does not mention it, but this falls into evolutionary game theory, since hawk vs. dove for finding stable evolutionary play. Sirlin's closing arguments are theory is no substitute playtesting. The author agrees with this due to the evolution of emergent behavior and decision making.

Sirlin's part three was all about fairness (Sirlin, Balancing Multiplayer Games, Part 3: Fairness, 2015). The primary discussion was about tier lists. Sirlin identified that he thinks the term originated from fighting games, in which there was a categorical ranking for each of the characters to express how strong or weak a character is based on an approximation. Sirlin used the term linear rank for characters, however the topics he discussed are different between ordinal rank; Sirlin's examples versus how magnitude strength between rank 1 and rank 2 is measured. Sirlin used a five-category tier list for his play testers and players and hoped that no character was placed in tier 1 (God tier) or tier 5 (Garbage tier). When all the available characters were in tiers 2 through 4, Sirlin stated that the game is in a balance. These tiers were created based on player perception, and no real quantitative approach is described in this article. Sirlin continued his article of his method of game balance by first removing the god tier characters and focused on the bottom tier characters until the character cast was close to 50%. Sirlin identified there are psychological effects to making a game rebalancing decision, however this paper will not focus more on the internal balancing of a game and will focus on the characters that are available.

Sirlin moved from expert opinion tier lists and began to discuss counter matches. Sirlin argued point that a strong character should not have matchups in which they are at a disadvantage. He pulled examples from other genres of games, and his argument rests solely on "But in Guile's case, you pick a character in a fighting game, you are stuck with him the entire game, so it really is a problem if he has some bad counter matches, even though players rate him fairly highly overall." The author missed exactly why having a counter match up in a fighting game is bad, it appears that it is an axiom that Sirlin has made for reasons.

Sirlin's fourth and last article about game balance is on intuition (Sirlin, Balancing Multiplayer Games, Part 4: Intuition, 2014). Early in the article he discussed how to know who the best players are. He proposed that the only way to do this is for all the players to play one another, no formal rating or number of games are mentioned, but it can be assumed a round robin format could be used. Sirlin then goes to explain that players may not be able to articulate in words their play style or experience, thus playing each other is the only way to know.

> "Why are the best players not necessarily able to reveal themselves as best through interviews or speaking? I claim there are two reasons:
>
> Spoken and written answers have extremely narrow bandwidth.
>
> It's impossible to access many of our own skills with conscious thought."

The discussion focused on why expert opinion is important for game balance. Sirlin utilized examples from Richard Feynman's approach to solve the case of the Challenger Space shuttle to his personal experience with a player of one of his games that solved sub problems. Sirlin's main discussion point was that expert opinion will be approximately right, and when time is a concern, expert opinion is preferred, especially when having to balance a game before a release date. Sirlin makes no mention to the continued review of a game for rebalancing and if this approach still stands.

Other articles by Sirlin explored his approach to game balance. Two articles specifically about Game Balance and Yomi and Game Balance and Fantasy Strike have reoccurring themes of matchup charts and qualitative assessment (Game Balance and Yomi, 2015). Sirlin referenced many tier lists and matchup charts from a variety of games. The game of Battlecon, a card driven game, is also referenced as a negative to not use quantitative data, since the data collected was small, including many matchups that only have a sample size of one, and the largest matchup having a sample size of 14. The remaining charts are expert created charts that will be analyzed in chapter 4. The tier lists also follow a standard of a uniform distribution that is used in determining a qualitative measurement of a character's strength. Sirlin placed emphasis on his card game Yomi and compared the matchup charts to other known fighting games that appeared in the EVO line up including Street Fighter 2, Street Fighter HD Remix, Street Fighter III: 3rd Strike, and Guilty Gear Accent Core. All of these are qualitative in nature and are not compared to any quantitative results. Attention should be brought to the evaluation that the uniform distribution is not a good indicator and attention by Sirlin suggesting correcting unfavorable matchups is the greatest importance for balancing rather than reducing tier spread.

Finally, Sirlin wrote about (Game Balance and Fantasy Strike, 2019), his latest fighting game as of this writing. In this report Sirlin again discussed only using qualitative approach to balancing, but in this publication discussed empirical evidence of matches for comparison. Sirlin also provided some anecdotal evidence of his own samples with the matchups for comparison and made the comment "When filtering the above table of expert player results for just games that I personally played, I found that my Jaina vs expert Geigers is 6.7 - 3.3 in Jaina's favor (compared to 3.4 - 6.6 the other way around). Admittedly a small sample size there, but I feel like it's a case where more Jaina players just need to step it up."

2.10. Fighter's Value

A research white paper was written to Berkeley's Student Association for Applied Statistics (Wang, 2019). Wang's research examined the fighting game Guilty Gear and focused on analyzing specific situations as it related to neutral position and corner positions for tracking on screen and in game decisions each player make for predicting outcomes. The research documented how visual sprites in game will be used and automatically catalogued for future analysis. The primary predictive model that Wang selected was based on Boice (2018). Boice's predictive model focused on scoring attempts and positions for soccer. Each of these positional attempts to score was assigned probabilities, and from this, a probability chart was formed to predict the number of goals that will be scored during a game. Boice's logic followed if they could predict the number of goals each team will score, they will be able to predict the results of a soccer match.

2.11. Rating and Matchup Predictions

An integral part of the style and pairing system comes down to the ranking system. From the initial rank to the adjustment in rank with games played, there is a need to have an accurate measurement of a player's skill level that is to be used for matchmaking pairing, bragging rights, and tournament entry. There are two main focuses of research in terms of why rankings are researched. The first is to predict outcomes of matches, and the second is toward betting margins.

First, rating systems should be based on how much luck is built into the game. Larkey et al. (1997) defined different variations of skill versus luck. On one extreme, there are pure luck games, followed by a mixture of luck and skill games, like poker, and finally pure skill games like chess. Fighting games would fall into the pure skill category for this paper as all actions deal the same desired effect. The only aspect of luck that is present is how the opponent reacts that enables it to follow many game theory practices as mixed strategies, however since each of the actions are decisions based on the players, the mixed strategies do create some luck factors that are observed in other sports like football or basketball.

McHale et al. (2012) investigated using past performances using wins and losses to determine a rank for FIFA teams. They found that using wins and losses was a typically poor indicator to be used as a forecasting tool. The authors then continued to attempt to build their own forecasting model for the FIFA league that would be a better indicator for forecasting a team's performance that used a variety of metrics such as manager, amount of star players, and rookies to determine the correlation that each of these individual metrics had on creating a team forecasting or absolute ranking.

While other indicators were predicated on individual indicators to determine a rank, other research looked at the relationship of the concept of quality wins and how time plays a factor for a player rank. The two games that were used as a base were MMA and tennis. "We only need to into account a recent sequence of match results for establishing a ranking of players, without considering the actual tournaments and a complex punctuation system" (Júnior, Gonçalves, Laender, Salles, & Figueiredo, 2012). The conclusion that in games in which it is one versus one, only recent performance should matter in determining a rank. The authors constructed a forecasting rating that used exponential smoothing to put a heavier weight on recent performance while older performance would play a smaller factor in the player's individual rank.

Football has been the focus of research to compare the efficiencies of the betting ratings versus the FIFA rankings. This research was led by Santos (2019) and found there is a gap between the two rating systems and found that with their developed rating system that implementing it compared to the betting rating there is a dominant strategy to earn a profit. The betting rating system can also be observed as an open-source rating system, a rating system that is based on an open field unweighted.

Two papers focused on the effects of a population of voters that can possibly have a negative effect on the rating. Guha (2003) wrote a professional publication that discussed the research at IBM that there can be voters that have a conflict of interest that can impact the effectiveness of the collective rankings. In short, there needs to be honest reviews to be able to have a fair and accurate ranking. Yang et al. (2009) researched how to battle dishonest behaviors in online rating systems. Yang et al. addressed that in an online environment there can be digital attacks or programed bots that produced skewed data voting mechanisms. The paper examined algorithms and situations that could identify when dishonest behavior was used in the rating system.

2.11.1. Player Hidden and Observable Rank will be Measured

Traditional sports have been a highlight, and each have used their season of play to determine the final tournament, there is a new wave of games that have come about that are still determining their seasons for their final tournament. These games are categorized as eSports. A unique characteristic between eSports and traditional sports is the number of games that can be played in a single day and the average game length is typically significantly shorter than most traditional games. Having the ability to play more games provides more historical data to determine player trends. The increase in data points provide the opportunity to gather a more measurements of a player's skill allow the analysis of a player's rank. What is expected is how to identify a large group of players to effectively sort them to determine the top 8 players for the final tournament.

A unique characteristic in many fighting game tournaments is the open tournament format. An open tournament is a tournament that has no prerequisites of skill to enter. This means that unranked players have the potential of facing off high-ranked players. The current tournament standard in fighting games is to use a double elimination tournament. The match making for this double elimination tournament use ordinal ranking of the known ranked players (highest rank receiving rank 1) and each proceeding rank given accordingly. When a double elimination bracket is made, the format that is used is the highest ranked player will face off the weakest player in the first round. This match making is used regularly even in traditional sport tournament as this methodology has the highest chance of providing the most exciting finals (measured as the gap between perceived skill of players).

The methodology for an 8-player tournament has fewer issues, but these issues become an increasing problem as more players are entered into the tournament. For example, looking at how March Madness basketball tournament is constructed, there are 32 initial teams, and the likelihood of rank 1 losing to the rank 32 team is extremely rare. The king-making tournament heavily favors to a strong bias degree that a repeated tournament game (i.e., a tournament played every year) will heavily favor a team that won or placed high in the preceding tournament for pairing purposes since the skill gap will be the largest for the next tournament. Compared to a single or double elimination tournament, other games have implemented a round robin or Swiss style of tournament play. Changing the number of opponents and the skill gap have impact on the number of games that are played, and the potential for more evenly matched games in most cases.

2.11.2. Performance vs. Skill

There is a difference between performance and skill. Performance is the results of a competition, while skill is the technical execution and decision that take place during a skill check. Duality of measurements are at odds with one another as skill is used to measure the potential. The purpose of this paper is strictly observing performance as our rating system. With repeated games, there are some assumptions that player skills increase in between tournaments, include hidden skill, and that the increase in skill effects performance. The purpose of this paper is not to identify specifically how much effort an individual player needs to put into potential to receive a certain performance rank.

2.11.3. Rating with Expectation

The two types of rankings that are of concern from a theoretic and practical perspective is ordinal ranking and performance ranking. Ordinal ranking is the ranking that is used for seeding. It is a rank of integer values starting at 1 for the highest rank player to the N value of how many players are entered into the tournament. These ordinal ranking is then used for pairings for the selected tournament. Ordinal ranking is easy to understand, which is why it is so widespread and easy to follow. It is with performance ranking that more calculations are required, and minor changes could have a big impact on how ordinal ranking will function.

2.12. Sabermetrics

Baseball is a sport that has tracked statistics and performance indicators since 1876 (Albert, 2010). During that time, simple calculations were used to evaluate players based on their hitting and throwing through means of earned run average (ERA) and hitting averages. These easy to measure and calculate performance indicators became a way of evaluating players. An on-going issue with certain quick and easy calculations is that the calculation may not provide any actual information on the performance of winning. Any figures can be divided by its sum, however what does that mean? As statistic techniques improved over the years, later these metrics were able to be compared with one another to observe how these hitting and pitching metrics are associated with wins.

This is where sabermetrics comes into play. Simply put, "sabermetrics is the science of learning about baseball through objective evidence" (Albert, 2010, p. 2). What this means is that baseball and other sports have many different types of data to capture, but finding the right question is the difficult part. To make informed decisions, it is important to analyze the correct measurements to answer the questions, and there are many pitfalls that can take place. For instance, the clutch metric to evaluate hitters in "clutch" situations, was found to not be correlated at all and was associated more with a luck factor. The clutch statistic helped create drama and spectacle but was not a great indicator for performance.

Albert (2010) described the development of key performance metrics for batting, pitching, and fielding. He also included the evaluation of creating scatter plots to identify correlation. The scatter plot of the OPS (on base percentage plus slugging) has a strong positive correlation to predict runs scored. OPS fair exceeded the traditional measurement of batting average that does not have a strong correlation to determining runs per game.

H – Hits

AB – At bats

BB – Base on balls (Walks)

HBP – Hit by pitch

SF – Sacrifice flies

1b – Single

2b – Double

3b – Triple

HR – Home run

OBP – On base percentage

SLG – Slugging

OPS – On base plus slugging

Batting Average

$$Batting\ Average = \frac{H}{AB} \tag{1}$$

On Base Percentage

$$OBP = \frac{H+BB+HBP}{AB+BB+HBP+SF} \tag{2}$$

Slugging Percentage

$$SLG = \frac{1B+2*2B+3*3B+4*HR}{AB} \tag{3}$$

On Base Plus Slugging Percentage

$$OPS = OBP + SLG \tag{4}$$

Below are the scatter plots for the 2008 baseball season from (Albert, 2010) research. The first scatter plot shows that batting average has a weak correlation, thus not a good indicator. While the second scatter plot uses the OPS stat that has a strong correlation, making it a good predictor indicator.

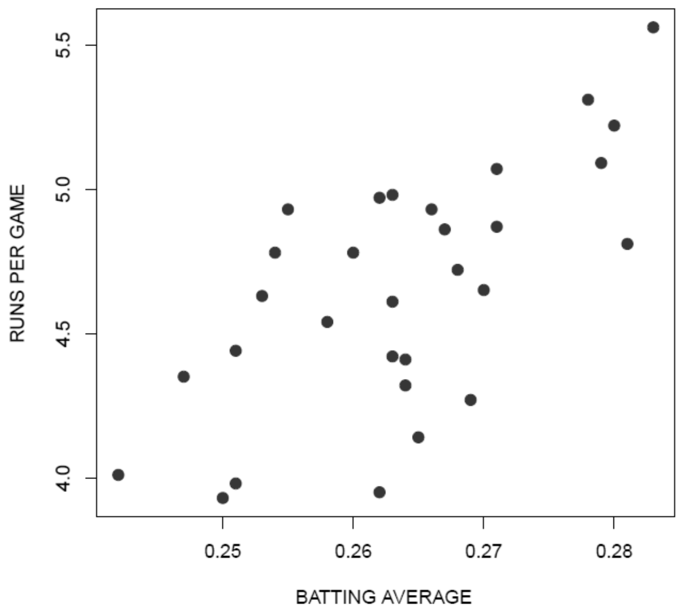

Figure 5. Batting Average vs Runs per Game
Figure adapted from (Albert, 2010)

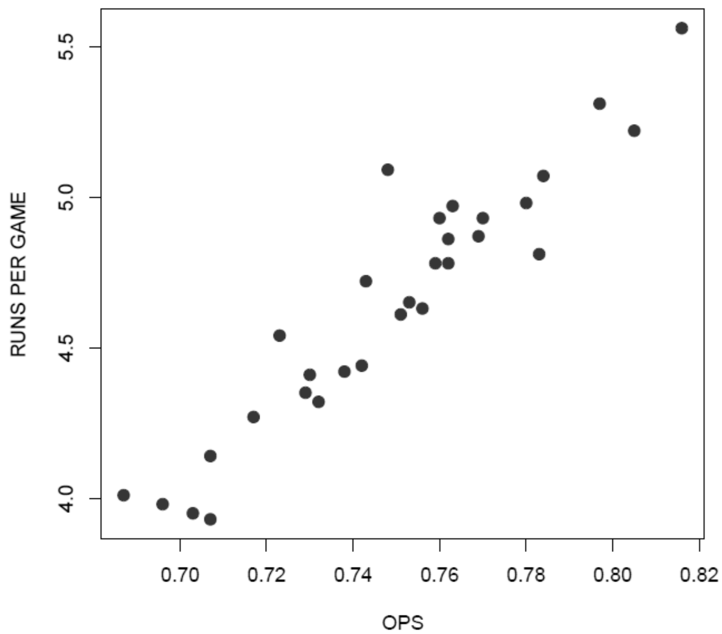

Figure 6. OBS vs Runs per Game

Figure adapted from (Albert, 2010)

Similar methods to analyzing hitting to find the correlated figure for run generation, is also used for pitching and defense metrics for fielding. Earned run average (ERA) is used to measure a pitcher's effectiveness at the plate, however, there are other factors that can have an impact on a pitcher's performance such as the defensive ability of the rest of the team. When comparing ERAs over time there is not a strong correlation that can be found with ERA itself. However, isolating different statistics for pitching performance there are more correlated figures such as strike outs. This finding sparked investigation into how to use an ERA like statistic to measure the predictive performance of a pitcher which lead to the defense-independent component ERA (DICE).

SO – Strike Out

IP – Innings pitched

Defense-Independent Component ERA (DICE)

$$DICE = 3.00 + \frac{13*HR+3(BB+HBP)-2*SO}{IP} \quad (5)$$

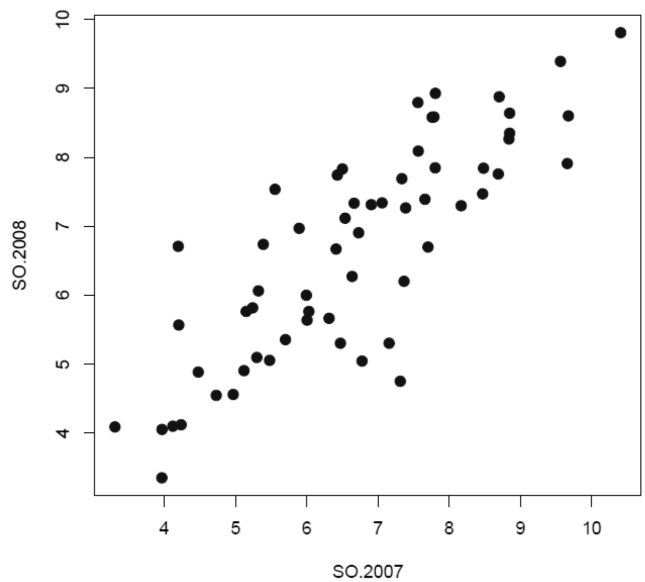

Figure 7. Strike Outs 2007 vs Strike Outs vs 2008

Figure adapted from (Albert, 2010)

From the above analysis and steps, analyzing the performance indicators for offensive and defense statistic differ, however the approach to running the numbers is the same. The more isolated a single statistic is, the more important it is to isolate that metric for use in the overall equation.

2.13. Baseball Statistics in Player Lineup Selection

Opposite hand advantage and in the overrepresentation of left-handed players in major league baseball was researched by Chu et al. (2016). Over 1.3 million data points were observed to identify left-handed hitter have about 15% opposite hand (OH) advantage for the on base plus slugging (OPS) statistic compared to right-handed hitters that have 7% advantage. Chu et al. also argued that even though left-handed hitters have an advantage to hitting in baseball, the equilibrium of left-handed players has not been met, and there can be an increase in the number of left-handed hitters to be part of the MLB. “Left-handed pitchers do not have to be as good as right-handers to earn a roster spot in the MLB” (Groothuis & Hill, 2018, p. 589). It was also observed that baseball batter selection followed a sequential game that allows counter picking common practice in baseball to make substitutions for either batter or pitcher to increase the chances of the substitution’s success. The paper emphasis the 5-stage game as follows:

> “Stage 1: The team manager determines the pitching rotation for the season. It is a practice in MLB that teams announce the starting pitchers for the next few games, so the pitching rotation is known in advance.
>
> Stage 2: The batting lineup is arranged to match up against the opponent’s pitchers to achieve the highest expected payoff.
>
> Stage 3: Starting batters confront starting pitchers in the matchup. The result of each confrontation follows the payoff function in Table 2.
>
> Stage 4: During the game, both teams substitute player to gain the highest expected payoff for the subsequent plays. Individual players are substituted due to injury, fatigue, ineffectiveness, or other tactical reasons.
>
> Stage 5: The outcome of the match is finalized, and the winner is decided.”

The methodology of the game theory equilibrium was to develop a starting line up with the proportion of left-handed and right-handed batters to maximize the outcome of winning the game. Groothuis and Hill identified some characteristics like experience, experience squared, OPS same handed, batter seasonal plate appearances, year, OH matchup for left hitter (LH) vs right pitcher (RP), LH vs left pitcher (LP), right hitter (RH) vs RP, RH vs LP, switch hitter (SH) vs RP, and SH vs LP. Groothuis and Hill performed correlation analysis of the above listed categories and yielded evidence of the “fighting hypothesis” (p. 1628) described by Raymond et al. (1996) which provided a theory of why the population is not 50/50 right and left-handed. Based on evolutionary needs of hunting with a dominant right hand, an equilibrium point of 90/10 for right and left respectfully was found was postulated by Raymond et al. Ultimately, analysis yielded LH vs RP has an OPS advantage of approximately 15% while RH vs LP has an advantage of 7%. Consideration of unfamiliarity of a specific matchup could help or hinder either the batter or the pitcher in this case, however no special consideration was made for style of pitches, i.e., local decisions in the matchup.

2.14. Prediction Performance

When it comes to formulating predictions, everyone wishes they held the all-knowing crystal ball. Knowing the performance outcome of a project, the stock market, or even a game becomes invaluable. Games have multiple methods to predict winners, but there are new rating systems that come out frequently with updates to improve the prediction. ELO is the most known and used, but the ELO system has known flaws. Glicko method attempted to solve the time element that is a weakness in the ELO system while maintaining many good elements (Glickman, 1995). The purpose of this research is to examine some popular used rating methods to determine the best method for pairings based on tournament style.

2.14.1. ELO System

The ELO system was developed by Arpad Elo in 1959 and adopted by the World Chess Federation in 1970 (ELO, 1978). The system was designed to predict the winner in a 1 vs 1 format by comparing player 1 rating (Ra) to player 2 rating (Rb). The algorithm is set up as follows:

$$E_i - Expected\ percentage\ chance\ of\ winning\ , i = A, B$$

ELO Expected percentage of winning for Player A

$$E_A = \frac{1}{1+10^{(R_B - R_A)/400}} \tag{6}$$

ELO Expected percentage of winning for Player B

$$E_B = \frac{1}{1+10^{(R_A - R_B)/400}} \tag{7}$$

ELO Expected winner check

$$E_A + E_B = 1 \tag{8}$$

$$S_i - \{1\ if\ win;\ 0\ if\ lose, ;0.5\ if\ draw\}, \qquad i = A, B$$

ELO Updated Ranking

$$R`_A = R_A + K(S_A - E_A) \tag{9}$$

The ELO system is a zero-sum rating system. This means that after a game has been played $E_A + E_B = 1$ which is the percentage of chance to win which is used to multiple by the k factor. The reasoning why the ELO system uses the zero-sum rating is because the true rating of each player is unknown so by using this Bayesian updating related to player's skill, the ELO system tries to predict the outcome. As a player continues to play games, it is assumed that the player's ELO rating will be closer to what their true ranking is. The ELO system also provides a predication of the outcome of the match. This is to say, a player that has a higher ELO rating is expected to win a game versus a lower ELO rated player.

Since the expected point values are an estimate, additional outcomes that eventually lead a closer approximation to a player's true rating. The ELO system uses a special k-factor in the equation to make large or small changes to the players ranking. There is some debate as to what an appropriate k-factor is, but what is used in chess is for lower players to have a k-factor of 32 while high ranked players use a k-factor of 24. This k-factor varied depending on the rating system they use, with the traditional rating used a k-factor of 10. Originally there was the assumption that the ELO rating system should follow a normal distribution, but later decided to use a logistic distribution which meant the staggered k-factor distribution.

There is also a performance rating period which for the ELO system it recommends n=20 games to gain understanding of a specific player's skill (Glickman, 1995). During this time, a performance rating metric is used. The performance rating can also be used for tournaments to find out a player's performance for that specific event, which may or not be used for their overall ranking. This algorithm is used as follows:

ELO Performance Rating

$$Performance\ Rating = \frac{[Total\ of\ Oppoents' ratings + 400 * (Wins - Losses)]}{Games} \tag{10}$$

2.15. Principal Component Analysis

Working with large data sets provided the opportunity to use a variety of different analysis techniques to assist with answering relevant questions. Player skill, opponent's character, stage, and side selection each reflect on a player's overall performance. Each of these indicators were recorded in combination and need to be separated to measure individual variables. The analysis technique that was used for isolating the variables and checking correlation analysis was principal component analysis (PCA). PCA is a statistical technique that looks to isolate variables and the analysis of a combination of variables.

PCA was first discussed by Pearson (1901) which laid the foundation for working with a matrix of data with the purpose of identifying correlated items in groupings. Pearson identified when analyzing independent variables in isolation frequently yields different results. However, combining independent variables into combination variables can explain the variance that occurs within a data set. One use of PCA is to reduce the data in a data set to identify the primary variables. PCA has been used in a variety of different research fields, and often under different names. Eigen vector and Eigen values are ways PCA is computed with identification of weights of variables and the planes to which the variable(s) demonstrate a specific weight to a data set (Abdi & Williams, 2010).

PCA by itself is a helpful tool, however one of the drawbacks is to requirement to provide a complete data set and deterministic data. Two variants of PCA were developed to handle these different situations with probabilistic principal component analysis and sparse probabilistic component analysis.

Probabilistic principal component analysis was first researched by Tipping and Bishop (1999), having identified the gap in which probability density functions are not used in the established PCA research. The variants to the main PCA model take into consideration the maximum-likelihood estimator for the respected σ^2 of the data set. Probabilistic principal component analysis still uses ortho linear arrangement for the analysis of the variables with attempting to identify the maximize variance of the data set. Eigen vectors are calculated to graph the planes with the respected variables and provide insight of the dataset of the weights of the reduced variables.

The final PCA subcategory of research of interest is sparse probabilistic principal component analysis. Up until this point, deterministic and probabilistic research have been mentioned, however there has been an assumption that all data is known to make a summary judgement. When data is incomplete, can these techniques still be utilized? The missing link was described in sparse probability PCA by (Guan & Dy, 2009). Their worked built from other researchers addressed the issues and situation of missing data and working with probabilities to identify the principal components of a data set. Guan and Dy tested their data with three real world examples with a constructed matrix of 960 by 624 and demonstrated the technique can be used with large data sets. Their contribution was the transpose matrix created before the computation of Eigen vectors to identify the principal components of the data set. In their example however, they discarded the missing data rather than replace it.

2.16. Summary

` The three levers of a tournament selection process have been outlined for the foundation of the proposed research. Understanding each of the pillars provided the framework of building models and measuring what is currently in practice. Limiting the selection process of each of the pillars to the unique scenarios of fighting game tournaments enable further research of this subject through narrowing the subject matter. The next section is Chapter 3 which will cover the methodology of the proposed research.

CHAPTER 3. METHODOLOGY

Research Approach and Hypothesis

- Collect data from previous played Fighting Game Tournaments from game footage
- Create a database of these recorded metrics for players and characters for each game
- Establish a Player Performance Report
- Utilize techniques to forecast expected tournament performance

3.1. Overview

The research was to investigate indicators for measuring Tekken 7 performance for players. Creating quantitative matchup charts, and predicting a players expected placement in a tournament were the topics of investigation. Answering these questions required analyzing historic performance tournaments and high-level play data to find trends and evaluate correlation of specific indicators. Previously recorded matches were observed, and documented tournament results found on smash.gg. Time series data was important to track trends and changes of time. The end results of individual games were analyzed to draw conclusions for pre-game decisions like character and stage selection. The variables that were tracked is found below.

- Date - The date is to use track the order in which tournaments took place to see trends in player and character data.
- Tournament Name - The Tournament name is to see if any specific tournament follows characteristics. Knowing the tournament should increase the chances of repeat players and possibly repeat characters that can be used to identify location groupings. It is also possible that certain tournaments or players are meta defining.

- Face to Face or Online – Tracked to differentiate between tournament formats as online play has a few rules broken when compared to face to face, such as each player can select side 1 and the online net code will automatically make each player perspective side 1.
- Tournament Format – denotes if the tournament is a double elimination, single elimination, round robin, or Swiss style tournament.
- Round # - running count of the round # for the specific tournament. The rounds tracker is also used to define total rounds and average rounds played.
- Match # - running count of the match in progress. Total count is used to measure total amount of matches recorded out of the max matches.
- Set # - tracks the count of sets played between 2 players.
- Player name on Side 1 – Player name that played on side 1.
- Player name on Side 2 – Player name that played on side 2.
- Game # - tracks the current game number that occurs in tournament order.
- Stage – The stage title that is selected, later used for categorical evaluation of what stages have preferences to character usage.
- Character selected on side 1 – a category that is analyzed to identify side preference to players and characters.
- Character Selected on side 2 – a category that is analyzed to identify side preference to players and characters.
- Player 1 win (Health Reaming) – Measured the amount of health remaining for calculating the damage ratio.
- Player 2 win (Health Reaming) – Measured the amount of health remaining for calculating the damage ratio.
- Last game win – A binary value to assist in terminal game.
- Last set win – A binary value to assist in terminal set.

- Last match win – A binary value to assist in terminal match.

The data collected was used in several different ways: 1) create a count of each character matchup that is played. Creating this count provided the data to measure win percentages for each of the character matchups. The matchup charts helped solve the sub-game that is associated with tournament matchups for game outcomes.

2) Have a running count of the sample of top players and tournament players measured character popularity. The count identified which sample gathered data representative of the population. The data collected was not disproportionate from one or two players, but for a collection of players using each character. The exception to this was when only one or two players used a given character. The popularity was a sample size of character selection and was compared to the Bayesian Nash equilibrium (Gibbons, 1992, p. 413) of the character matchup chart created.

3) Document players win/losses in each tournament. Using prior performance based on players the predictors of future player matchups was created. This information assisted the prediction methodology development to determine tournament placement based on the players who entered the tournament. Damage dealt and damage remaining were indicators to track magnitude of win/loss. The quality of win/loss was expected to have a high correlation to determine the victor of a match as it is used in other traditional sports with respect to runs scored and runs allowed.

3.2. Tournament Game Extensive Form

Tournament performance were observed to document the decisions that occur during a tournament, including character pool creation and stage selection. The extensive form tournament structure in Figure 1 was used to analyze the decision points to measure the effects on each player's probability of winning. The extensive form summary steps that take place for each tournament match is as follows:

1. Side Selection
2. Character Selection I
3. Game 1
4. Character Selection II
5. Game 2
6. Character Selection III (if tied after Game 2)
7. Game 3

This extensive game created a 116-node tree of all the possible outcomes for a single tournament match. The primary decision points analyzed were the subgames that occurred. Side selection, character selection 1 through 3 provided the subgame decisions, while the outcome of games 1 through 3 were the performance metrics.

3.2.1. Side Selection Subgame

The first decision in a fighting game tournament was side selection, the preference of the player to sit on a preferred side of the table. A global overview of win rates for both sides to identify if there was a side that has a global advantage over the other, rather than based on pure preference. The outcome of side selection was completed in one of two ways.

The subgame began with the players agreeing to what side to sit on. One side selection strategy was through preference selection. Either both players agreed to which-side they played on, or the first player selected, and the opponent made no objection.

The second way was to request a rock-paper-scissor (RPS) match. This created a quick subgame in which both players had equal chance of selecting their preferred side, but it is possible the payoff schedule was different. The winner of the rock-paper-scissor match, typically best of one, but occasionally best of three, the winner of RPS selected the side they prefer.

3.2.2. Characters Selection Subgames

Unless tournament rules specify, there were no restrictions to character selection, and it is possible that the same character was selected by both players, called a mirror-match. After character selection, the characters had a limited number of moves including a combination of movement, attacks, and throws to eliminate the opponent. As is common in fighting games, some characters are considered stronger than other characters based on specific characteristics related to damage output for moves, speed, and/or lacking a type of move all together. It is possible for a player to select a character that has strengths compared to a different character but could be considered weaknesses to others. In practice, these are referred to as matchup charts.

3.2.3. Character Selection I Subgame

There were two processes of a character selection subgame. The first process was both players freely sat down, and one player waited until their opponent selected a character, then the second player immediately counter picked a character. A counter pick means that a there was a second mover advantage. The matchup charts created helped aid in the decision process for counter pick selection.

The second process of the subgame was for a-player to request a double-blind pick. In a double-blind pick, one player told a judge which character they selected, and after the character was known by the judge then the opponent selected a character on screen. The judge then announced the character chosen and the first game began (Evo Tournament Rules, 2020).

The proposed solution to this subgame was solved using Bayesian Nash equilibriums with data from the matchup charts regarding character selection probability of being selected. For instance, it was assumed some characters have a lower chance of winning if allowed an initial counter pick, so it was in that player’s interest to request a double-blind pick in hopes of having a favorable matchup in game 1, thereby have the option to counter pick before game 3.

3.2.4. Character Selection II Subgame

Subsequent iterations of the character select subgame occurred before the second game and possibly before the third game. This character selection gave a player the option to do one of two options. Option 1 was to keep their losing character but change the stage (the on-screen location with set parameters). Option 2 was to change their character, and a stage be re-randomized. Depending on the game, some options were limited. The solution to this subgame was to solve three separate subgames and select the option that is best with the added information that the winning player cannot change their character.

3.2.5. Character Selection III Subgame

Character selection III subgame was nearly identical to character selection II subgame that is described in the previous section. The only difference was from the game 2 losing player that should make a character selection to maximize their chances of winning without having to consider the game 2 winning player’s future response, as there was not a game 4.

3.2.6. Games

The game play itself took place with the characters and stages that were selected in the previous sections. The players competed with 60 seconds per rounds to either deal damage greater than or equal to the opposing players starting health or have more health than their opponent at the end of 60 seconds. There were many factors that contribute to the offensive and defensive attributes and the decisions make fighting games so exciting. For the purposes of this research, individual in game decisions were not tracked. Future research should include the investigation of learning purposes of a fighting game, but this research focused is on high level player results and character assignment.

3.3. Matchup Charts

Matchup charts show the percentage chance of a specific combination of characters in which one character had an advantage to win. This is due to the tools characters have during a game. The game of Go and chess have similar advantages in terms of first mover advantage. To eliminate the first player advantage, Go instituted a rule to combat the first mover advantage by requiring the first player to start with a point penalty. Historically, Go did not have today's statistical measurements for many years and was unable to track or implement a new standard to eliminate the first moved advantage. Go was played as an even game, with no komi to balance the strength of having the first move. Over time, players and tournament organizers observed that the first player was disproportionately winning, thus the added komi. The handicap has adjusted over time, and even depending on the area in which the game is played, the handicap during the same year in Japan was different than the komi that was used in Korea for instance. The game of Go analyzed historic games at a high level and imposed the handicap to eliminate the first mover advantage. No academic sources of the history of komi could be found, only rule updates from tournaments and a brief discussion at https://senseis.xmp.net/?Komi.

Chess on the other hand observed the first player advantage, but there is no handicap. While in Go there is always one winner and one loser, in the game of chess, the game allows for a draw match to occur. Both games, since these are sequential games, both have the same characteristics of first mover advantage. Each of these games address this in tournament play for playing a first to two or first to three and giving the previous champion or higher ranked player this supposed advantage, thus when the underdog wins, it is truly because of skill.

Fighting games differ as they are simultaneous games, and the players have the option of selecting their character before a game. This creates a subgame before the match begins and one decision aid that exists for players during character selection with the use of a matchup chart. The creation of matchup charts is not to show that each individual character is balanced, but to show that the roster itself has an equilibrium. During the discovery of recorded matchup win probabilities between characters, the evaluation of the created charts provides information of a character being strictly dominated when compared to other character selections, then that provides information to the player base for character selection.

The round tracker was used to measure the win/loss probabilities for each character pairing for the matchup chart. Counts for rounds, games, and sets were tracked to determine the historic win percentage for each of the pairing with the calculation of win percentage as $Win\% = \frac{Total\ \#\ of\ Wins}{Total\ \#\ of\ Games}$. This calculation was placed in a symmetric matrix as in example in Table 3. A minimum target of 30 recorded games was imposed to provide some statistical significance with each character pairing.

After the matchup chart was created, Bayesian Nash equilibrium was used to find the equilibrium of characters that were selected. The developed matchup chart was a great improvement over the voting style and other qualitative measurements that were in use, to the extent that top players have zero confidence in tier lists and matchup charts because of the high qualitative nature. This will put all those arguments to rest as the adaptive method described here to address major issues and the matchup chart will be able to be used for decision making for practical use, rather than just for fun.

3.3.1. Matchup Validity

As data was collected for each matchup using the game form, the data was tallied in the order it which it was played (by date, then by round). The data was organized for each matchup combination and basic statistics run. The data was organized as a series of 1s for wins and 0s for loses to find a mean and confidence interval to determine confidence level. The mean represented the win rate of a specific character matchup. It would be ideal to have 400 recorded matches for a confidence interval of 95% and 300 matches for a confidence interval of 90% as each of those intervals will be less than 5%, thus it would feel confident of that matchup favor.

A factor that will have to be considered is the quality of matchup that are measured for this determination. To factor this in, the higher ranked player matches will be used. It should also be considered, if tournament data is being taken, a specific player will have more recorded matches for the matchup charts. In other words, due to the nature of double elimination tournaments, the quality of data gathered automatically filter the best players for a character.

When interpreting the data, keep in mind that each game is assumed independent of one another. If a character was switched between a set, the expected win percentage will be used after the character switch.

3.3.2. Character Selection

The matchup chart gave valuable information for analysis to understanding tournament format. Bayesian Nash equilibriums were used to determine the best permutation of characters for character selection to analyze which character to use or to practice against. The Nash equilibriums showed a percentage of use of characters under different conditions and the proportion of characters that were expected to be used in a tournament, a player was enabled to make decisions to increase their expected value with a pairing for a tournament.

3.3.3. Character Pool

Matchup chart analysis helped to reduce the number of characters to be used in tournaments. The reduction in the number provided the information of what characters to select from, and what characters to practice against. The reduced character pool space identified the highest probability selection including the dominate characters of the current matchup charts. The reduction of character pools eases the burden of the players to practice and memorize to much information, and strategically identified where time should be allocated to maximize a player’s probability of winning.

3.4. Predicting Tournament Placement

Any good predictive measurement has a most likely term and a range of probabilistic outcomes. Using ELO, character usage, and other players that entered a tournament, an estimate of placement was calculated. Double elimination tournaments are the norm for fighting game tournaments, double elimination rules are used.

The first step in estimating success required creating two categories of players. The nature of double elimination tournaments gives the higher ranked players an advantage in the tournament. Since that consisted of half of the players, that created the first two categories, advantage players and disadvantage players. This initial pairing showed that in the first round, the advantage player had a higher chance of winning due to pairing alone.

Double elimination tournaments have three different types of pairing: King making, McMahon, and random. King making pairs the highest rank with the lowest rank at each stage of the tournament bracket. This was designed for the top ranked players have an easier road to the finals and should in theory provide the most exciting games for the finals. The king making system shows that each round will have at least one potential promotion demotion match, i.e., a match in which both players are close to evenly match and has a chance of an upset. The McMahon system, as discussed in section 2.4., has pairings be progressively closer to one another. In theory the finals should like very similar to the King making system and making rounds in the tournament progressively closer. The last is randomly assign players to advantage players. Random assignment is typically not recommended but can occur if there are a lot of players that have no previous known rating that the tournament organizers are using as a metric. Randomly assigning players has random number of effects of close matchups and uneven matchups. This pairing should be avoided.

The prediction is shown in Table 9. The expected value of a players ending result is by the sum product of the probabilities of a result by the number of wins. The table is created from 0, representing a player receiving 0 wins to $2^{n-1} < N \leq 2^{n}$ 3n-1 (Table 6) which is the minimum amount rounds that were played. Since the loser side had twice as many games, wins in the winner bracket on the loser side were measure by 0.5 wins, and wins in the winner side are measured as 1 win.

Initial categories were advantage and disadvantage, but as the number of players (N) increase in a tournament, these categories became less reliable as buckets. That is why after analysis of tournaments, as N increases, more categories increased. At the time of this writing, it is unknown how to split and categorize these subcategories without data analysis, but it is one of the areas of the research that will be investigated.

Table 9. Expected Player Ending Placement

# of Wins	Probability %
0	
0.5	
1	
…	
$2^{n-1} < N \leq 2^n$ 3n-1	

Taking everything into consideration, five potential factors impact predicting tournament results. From the three previous metrics of player rating, stage, side, characters, and other players into the tournament. The categories will take the robust approach of historic games and other players to determine how to create the advantage and disadvantage categorize individual ranking.

3.5. Tournament Style

The tournament format that is the focus of the paper is to use the double elimination and single round robin. These tournaments are the standard for play for open tournaments and premier tournaments. Each of the styles use a best of three format for matches, which mean there retains the same extensive form for tournament play, allowing character switches. There are no current published methods that are used for the double elimination tournaments, that is why three different methods for pairing were taken into consideration. The king making method, McMahon, and random assigned were used in the analysis. It is the intent to gain insight of how tournament organizers pair players for initial tournament set up.

3.6. Rating

The rating methods follow two aspects. Analyzing the historic data of on smash.gg of the major tournaments, a best fit model analyzed players that have played in multiple tournaments to measure the end ranking and the number of players in the tournament. A best guess created using the ELO rating system ranked players following ELO standards to determine grand master, master, etc. rank of players. The second rating system analyzed of the combination of style and pairing to measure the accuracy of tournaments for placement matches and ending rank. A hidden rank, or true rank, was-assigned to players and as the simulated tournament ran, a measurement of error determined if a double elimination tournament accurate depicted the buckets of ranks created. The double elimination ending results were compared with the modified round robin play to see if a more accurate reading of a players hidden rank aligned with the ending position. For the hidden and ending number the ELO system was maintained as it is easy to use, and the potential issues of the ELO system as is related to time was assumed not be an issue. This leaves room for future research to determine a new rating system to better reward and accurately measure players.

3.7. ELO Integration Steps

1. Identify Rating System (ELO)
2. Create Matchup Charts (Historic Data)
3. Combine (Weighted)
4. Combine with Counter Pick

Rating systems emerged in games to assign a numerical representation of strength to players to assist with the objectives of matchmaking and observable rating. Potential issues arose when game developers started to observe different categories of players that do not play and look for rating incentives. These incentives come in the form of point inflation at lower ranks, that ultimately show a large disproportionate of higher ranks achieved, even though these match up may not be even. Other external factors such as geographical distance and internet latency before a bigger issue for the enjoyment of the game. While the enjoyment of the game is important, if there are common mismatches to a fair matchup, it causes further issues related to having a reliable rating. Game developers constantly evaluate all the experiences not just for tournament players, but for players of all levels. This phenomenon is also observed in other games like Magic the Gathering, a card game, that established their own rating and rank system with the objective to encourage players of all levels to continue to play. Magic the Gathering has its own issues and criticisms from high rated players, especially whenever there is a change to the rating system since player's behaviors change to maximize their own points for financial gain and/or position to increase their chances of winning a tournament.

An assumption that can be made is to have a "best fit for all" method, but for the meantime, the focus on establishing tournament rating and character matchups. Using established system, the ELO system was easiest to implement. An early concern is the minimum number of 30 games recorded before the ELO accurately depicts a player's individual rating. Some criticism of the ELO is that while it is theorized to be a normal distribution, empirical evidence shows that it more closely follows a lognormal distribution instead. These factors were considered moving forward. Another observation taken into consideration is that the rating assumes there is no other factor that gives a player an advantage before the match begins, other than one’s skill level. For fighting games, since the assumption is that each character has their own strengths and weaknesses, certain pairings were made in which one character has a distinct advantage over the other, but if two players with equal skill play each other over and over, it steadily increased one ELO to be equal to the matchup percentage, and being a zero-sum game, the other disadvantaged player had the ELO decreased equal to the matchup percentage. In other words, ELO assumes that the character selected is to have a 50% chance of winning, no advantage of disadvantage. Fighting game terms refer to this as a mirror match.

Phase 2 was analysis of the matchup charts. As discussed, the ELO does not figure out the advantage of a character, thus the ELO can be an unfair metric if used without considering the matchup charts.

Phase 1 and 2 worked under different assumptions. Skill rating assumes that the characters are equal, while matchup charts assume that player skill is equal to be true. The next step to bridge the gap of these assumptions required determining how the combination of a skill system and matchups accurately predict outcomes. A weighted approach was used to determine which of the two sides represent a better proportion of the weighted rating.

ELO Weighted Player Rating with adjustment Matchup

$$ELO * \alpha + Matchup * (1 - \alpha) = Player\ Rating \tag{11}$$

A topic that has not been fully explored, that may play a part, is the inclusion in character floors and ceiling. What is meant by that is how easy or difficult it is to effectively perform the character’s skills at different skill levels. Therefore, many pros discount some matchup charts source data and rely on top player expert opinion.

The final step was to observe how a player rating can increase by having multiple characters that are used for counter picking, i.e., to increase their matchup for a specific character. Information of each individual player's picks for an entire cast is relatively unknown, however, tracking previous games and capturing the matchup percentage, and increase the player rating in theory should increase the player rating. General sense, if a matchup has a higher chance of winning, but a player loses the match, the player rating of the player will decrease more as they had a perceived advantage. Conversely, if the advantage matchup had won, the increase to player rating would be smaller, and the disadvantage player would also only lose a small rating. An equilibrium should be met under both conditions to be used as a baseline for each phase.

3.7.1. Phase 1 – ELO assume Characters are equal (100% ELO)

Using the ELO formula to predict the expected outcomes required the ELO rating of each player. In a tournament setting, the effects of player rating are multiplied due to playing multiple sets. For example, a 1200 ELO rated player versus a 1500 ELO rated player was observed to have an 84.9% chance of winning. The same rating disparity in a first-to-two tournament, the win percentage changes from 84.9% to 93.8% of winning the match up (Table 11). This computation again assumed the characters are equal.

Table 10. ELO(P1) vs ELO(P2) = 1200 vs 1200

Game 1	Game 2	Game 3	Percentage
0.500	0.500		0.25
0.500	0.500	0.500	0.125
0.500	0.500	0.500	0.125
0.500	0.500	0.500	0.125
0.500	0.500	0.500	0.125
0.500	0.500		0.25

Table 11. ELO(P1) vs ELO(P2) = 1500 vs 1200

Game 1	Game 2	Game 3	Percentage
0.151	0.151		0.023
0.151	0.849	0.151	0.019
0.849	0.151	0.151	0.019
0.151	0.849	0.849	0.109
0.849	0.151	0.849	0.109
0.849	0.849		0.721

3.7.2. Phase 2 – Matchup assume ELO are equal (100% Matchup)

When matchup charts are utilized, a character advantage may have an impact on the overall expected win%. For the example in Table 12, assume the Player 2 selected a character matchup that should win 65% of games versus Player 1's selected character. It was observed due to the tournament format, a bad matchup makes the error multiplicative in that in a first-to-two Player 2 had a win rate of 71.8% instead of 65% with a single game.

Table 12. Character(P1) > Character(P2) = P1(0.65) > P2(0.35)

Game 1	Game 2	Game 3	Percentage
0.350	0.350		0.123
0.350	0.650	0.350	0.080
0.650	0.350	0.350	0.080
0.350	0.650	0.650	0.148
0.650	0.350	0.650	0.148
0.650	0.650		0.423

3.7.3. Phase 3 – Weighted Computation

Taking the previous two examples, the formula combining the ELO and matchup and assigning 50% to each ELO and matchup. With a 1500 ELO and 65% matchup for Player 2 showed a compounded advantage when compared to Player 1’s 1200 ELO and 35% matchup. The above chart showed the combination and the observed win rate for Player 2 went to 84.3% since the value was between the win rate for skill and matchup. This was not intended or expected, as it is expected that the higher rated player that picks an advantage matchup increased their win rate.

Table 13. ELO(P1)|Matchup(P1) > ELO(P2)|Matchup(P2) – 50/50% weighted

Game 1	Game 2	Game 3	Percentage
0.250	0.250		0.063
0.250	0.750	0.250	0.047
0.750	0.250	0.250	0.047
0.250	0.750	0.750	0.141
0.750	0.250	0.750	0.141
0.750	0.750		0.562

To consider the additive nature of a positive matchup, there needed to be a level figure and an additional figure. In this case, the ELO was our level that adjusted to incorporate a specific matchup. To have this adjustment a formula was introduced that took into consideration a specific player's historic performance for a matchup with the global perspective of a specific matchup.

Adjustment to the ELO used as a base and the match up adjustment added on. Starting with the ELO expected win percentage for a player-versus-player matchup, the matchup chart recalled solving for the rating adjustment. When solved for n in this scenario in the ELO expected win from equation 6, was simplified to:

ELO Matchup rearranged for Matchup

$$-1 * \left(\frac{ln\left(\frac{1}{MU\%}-1\right)}{ln\,10}\right) * 400 = n = MAdj \tag{12}$$

Using this equation enabled us to input a matchup percentage (MU%) to approximate the difference between the rank of player 1 and of player 2. For instance, if there is a MU% equal to 60%, the ELO adjustment will be approximately 70. The 70 can be placed all on player 1, or 35 can be placed on player 1 and 35 can be placed on player 2, the result of ELO expected performance will remain the same.

How this is used is to start with the ELO and incorporate the matchup adjustment as an addition.

Adjusted Performance Rating with Matchup Adjustment

$$ELO + MAdj = Adjusted\ Performance\ Rating \tag{13}$$

From the above example of player 1 with an ELO of 1500 and using a character with a 65%-win rate. Using the MU% equal to 65% returns a MAdj equal to ~107. Keeping the example simple, the 107 will be added to player 1's ELO for an adjusted performance rating equal to 1607 vs the 1200 ELO of player 2. The follow tournament expected chart is as follows.

Table 14. Expected Win Percentage 1607 vs 1200 ELO

Game 1	Game 2	Game 3	Percentage
0.087	0.087		0.008
0.087	0.913	0.087	0.007
0.913	0.087	0.087	0.007
0.087	0.913	0.913	0.073
0.913	0.087	0.913	0.073
0.913	0.913		0.832

With the Matchup Adjustment, the 1500 vs 1200 ELO increased from 84.9% chance of winning to a 97.8% chance of winning. Conversely, if these two players were friends and looking to have a fair fight, the 65% could be in player 2 favor. While the player skill is unchanged, and not accepting to deliberately play poorly, next is to compute the scenario of the scenario of a character switch in the disadvantage player's position.

Table 15. Disadvantage Player Character Switch 1500 vs 1200 ELO

Game 1	Game 2	Game 3	Percentage
0.248	0.248		0.062
0.248	0.752	0.248	0.046
0.752	0.248	0.248	0.046
0.248	0.752	0.752	0.140
0.752	0.248	0.752	0.140
0.752	0.752		0.565

Searching for a fair fight, player 1 went from a full advantage best of 3 with a probability of winning at 97.8%, down to a probability of 84.6%. Starting for the non-adjusted probability of winning of 93.8%. This does make it a better fight, but still standing at 84.6% is clearly still in player 1's favor, but now player 2 will pull off potentially more wins in long extended play sessions, since the per game win rate dropped from 84.9% down to 75.2%, however a very poor chance in a tournament showing.

3.7.4. Phase 4 – Counter Picks

The final phase was to incorporate counter picks. Counter picks refer to the process of selecting a character that has a positive matchup percentage. So far, the analysis demonstrated examples of a single character per player. Phase 3 compared the two possibilities of a high rated player selecting an advantage character and the lower ranked player selecting the advantage character for the overall probabilistic results for a best of 3. Now phase 4 incorporated the multistage game of identifying character chains and selecting the character~~s~~. For this section it does not need to be illustrated at this time for each ELO adjustment based on matchup percentages, but it can be identified purely on the matchup chart and character selection itself.

3.8. Player Performance vs Global Performance

Up to this point, the discussion has involved the analysis of global matchup win percentages without regard to personal matchup win percentages. This causes some player to immediately discount the use of global matchup win percentages, since their personal win percentage they feel is lower or higher than that, or that it does not matter at all. To assist with this argument is the example in baseball for the pitcher versus batter matchups and how handedness reflects an advantage or disadvantage in that matchup for pinch hitting or pinch pitching.

Research has used simplistic skill adjustment for making decisions to maximize expected outcomes, like a baseball manager selecting their batting lineup with the knowledge of the pitcher. For fighting games, each player has their own matchup win rate % that adds to the global win rate, however, since the ELO measured skill that is adjusted with MU%, the player performance needed to be considered along with the global MU%. The methodology used to track the global MU% for players considered what characters to play, and found the character equilibrium, however personal record effected the decisions of a matchup of player-versus-player.

Global Matchup Win Percentage

$$GMU\% = \frac{\sum_{i=1}^{N} Wins_i}{\sum_{i=1}^{N} Games_i} \tag{14}$$

Personal Matchup Win Percentage

$$PMU\% = \frac{\sum_i Wins_i}{\sum_i Games_i} \tag{15}$$

Personal Matchup Win Percentage converge to Global Weighted

$$MU\% = \left(\frac{M-n}{M}\right) * GMU\% + \left(1 - \frac{M-n}{M}\right) * PMU\% \tag{16}$$

N = Total Number of games that matchup is used

n = Number of games player i played that character

M = A set constant in which data points converge

If a character has previous knowledge, the global matchup chart served as an indicator of how the matchup will take place, and given enough data points, the personal average determined a specific matchup. Part of the analysis analyzed test data where a theoretic MU% is addressed and games were simulated with that win percentage only to record win or a loss. Standard deviation was analyzed to determine when the data series converges to an acceptable tolerance. Having a known number of games for convergence determined the value that will be used for M in equation 19.

A possible situation can occur that a PMU% is very different from the GMU%. If the PMU% > GMU% a possibility is the ELO mismatch occurred with that matchup. The opposite may also be true when PMU% < GMU% that the ELO mismatch occurred playing against higher ELO opponents to affect this number. Prior to a game, the PMU% of each player was compared to assess the average of the two players. A strict average was taken since these are the only two players in the match, however their adjusted matchup affected each other. When analyzing this comparison, the ELOs of each player was estimated, and the character matchup used for further comparisons as described in section 3.8.3.

3.9. Regression of Players

Players began to accumulate games as more tournaments occur. As games were recorded for tournament players, a scatter plot was formed for players with the x-axis is game and y-axis win percentage. As players were added to the scatter plot, least squares fit test was used to determine if specific matchups remained true, and to identify players with higher performance scores were identified. The least squares method helped determine the significance of the data set. This focuses on ending tournament placement and relative ranking based on the total number of players that enrolled in the tournament and the final placement of the players. If a player consistently ranked in a relational rank, it was concluded they were a consistent player, or if the player was on an upward or downward trend.

3.10. Tournament Results

Tournament results for each player were reported and tracked as a scatter plot. Players were expected to follow a logistic curve in which top players clumped together. This scatter plot has the x-axis as number of tournaments and the y-axis to be the relative rank of finishes of tournaments. This chart showed a representation of all fighting games. Individual games were tracked with the following characteristics.

- Ordinal Rank – Ordinal
- Tournament win points (TWP) – 1 point for winner bracket, 0.5 points for loser bracket
- ELO – Their current estimated ELO
- Character(s) used – What characters were used in the tournament
- % Rank – The amount of TWP earned divided by Max TWP

3.11. Player Report

Data gathered was ultimately used for player and character evaluation. Utilizing the raw data gathered, computations were used to find player key performance indicators to evaluate a player's strength based on previous performance. The following metrics are the initial player report that was generated:

- Games Played – Total number of recorded games of that player.
- Wins – Total number of recorded set wins.
- Losses – Total number of recorded set losses.
- Best Matchup – What character the player has the highest win rate percentage.
- Worst Matchup – What character the player has the lowest win rate percentage.
- Main – Most player Character as a proportion of all recorded games.

- Secondary – Second most player character of all recorded games.
- Reserve – Additional character used of all games recorded.
- Best Stage – Stage that has the highest win rate percentage.
- Worst Stage – Stage that has the lowest win rate percentage.
- Favored Side – Right or Left and win rate percentage.
- Unfavored Side – Right or Left and win rate percentage.
- Damage Dealt – Total Damage Dealt.
- Damage Received – Total Damage Received.
- Rounds Scored – Total Rounds won.
- Rounds Allowed – Total Rounds Lost.
- Games Won – Total games won.
- Games Lost – Total games lost.
- Damage Ratio – see below.
- Round Ratio – see below.
- Game Ratio – see below.
- Set Ratio – see below.

Table 16. Player Summary Sheet

	Player Season	Player Life	Season	Life Time
Player Name				
Games Played				
Wins				
Losses				
Best Matchup				
Worst Matchup				
Main				
Secondary				
Reserve				
Best Stage				
Worst Stage				
Favored Side				
Unfavored Side				
Damage Dealt				
Damage Received				
Rounds Scored				
Rounds Allowed				
Games Win				
Games Loss				
Damage Ratio				
Round Ratio				
Game Ratio				
Set Ratio				

The four columns provided comparison. The player columns represent the player specific statistics for the season and the total lifetime. The season column measured for the game season, since EVO was used as the primary tournament, while all seasons began and end with EVO. The lifetime column measured the games lifetime metric.

The ratios, a best fit using regression was used to determine what the constants and n factors will be. The same methodology was used for the predication of pointed scored and points allowed in other sports like baseball, hockey, and basketball (Rothman, 2014). These sports used sabermetrics, as discussed in section 2.11, in this evaluation to predict outcome of games using qualitative data, which is more than measuring wins and losses. The two equations that are used are as follows:

Win Rate using Runs Scored and Runs Allowed

$$win\ rate = \frac{runs\ scored^n}{runs\ scored^n + runs\ allowed^n} \tag{17}$$

Calculation of adjustment Metric for Global Win Rates

$$n = k * log\frac{runs\ scored + runs\ allowed}{Games} + b \tag{18}$$

k – constant smoothing multiplier

b – smoothing constant

In addition to player sheets, further computations can be used to evaluate a players win rates and game break down for the typical tournament structure of best of 3 format. To do this, individual breakdown of each permutation of wins and losses that can occur. These metrics can be used for tournament commentary during a game to build up expectations for the viewer based on the state of the game.

3.12. Measure for Success

The element for success is to determine the winner of a tournament as well as approximation of each individual match. Currently on https://liquipedia.net/fighters/All_Tournaments are databases of previously played fighting game tournaments are available. The site does not go further to estimate a rank that is public but only shows results of each player in their tournament, along with the opponents that they competed against. Purposes of this paper included the identification of upsets as an error metric in forecasting and ranking of players expected performance. Reviewing historical data created a base line and a model was constructed to develop a better estimate of the players expected ending positions, a target to achieve. It is assumed that this target is a better incentive for players if this number is public, since it is assumed, and only under very rare cases, that a new player comes and wins a tournament.

3.13. Summary

Chapter 3 developed the logic that was used for the core research of this report. After recording player and match data, data was analyzed to identify performance metrics. ELO rating system was used as a baseline, along with methodology from sabermetrics for isolated performance metrics. These metrics were then used to summarize a tournament player summary used for sponsors, teams, and tournament seeding in the future. The next step was the forecasting methodology from sabermetrics as well as historical placement of the players in the respective games to identify trends in player ranking tendency.

CHAPTER 4. PRESENTATION OF DATA

The analysis of data is gathered from observed Tekken 7 matches that had archived on YouTube. After the data had been gathered, game theory concepts using both Nash equilibrium and extensive form games were used to gain insight into decision making and predictive analytics. Game theory assisted players to analyze tournament subgames to make the best response based on opponent decisions. The subgame solutions are influenced with the method of backward induction that determined counter picks in the tournament subgame and stage selection. The predictive analytics that were used pertain to the solutions to subgames and verified using linear regression to calculate r-squared metric to identify variables that play the largest role. Principal component analysis was set up, however too much data is missing to accurate isolate variables and their combinations for further review. Next section is the presentation of the data that is converted into information to be used for the conclusions.

4.1. Overview

Data has been gathered from a variety of tournaments that were saved and archived on youtube.com. The list of tournaments came from Bandai Namco’s Tekken World Tour website (More fun for everyone!, n.d.), and tournaments that took place around the world for 2018 to 2019. Evo tournaments were also observed as the main tournament series as the Evo tournament have the highest amount of individual registered participants.

The primary task was to answer what were the decisions needed to maximize winning set % involved in the subgames side selection, character selection, and stage selection. A review of the research question of this research is as follows:

- How to predict outcomes of a tournament set?

- How is performance rating calculated for individual players?
- What decisions should be made to maximize probability of winning through tournament subgames? What are the strategies that are used for side selection and character selection with respect to the opponent?

4.2. Tekken 7 Tournaments

The Tekken World Tour (TWT) for Tekken 7 began in 2016 in NA. The event took place at the Final Round tournament series that took place on March 25, 2016. Tekken 7 and the TWT continued with the support of Namco, the game publisher, to provide the opportunity for competitors to market the game to a general audience. Many different tournament organizers structured their own tournaments while abiding by the rules established by the TWT. Tekken 7 had continued growth, especially at Evo from 2016 to 2019. Table 17 lists all the Tekken 7 tournaments that were observed for this research.

The announcement of the 2018 TWT official season increased the number of tournaments which occurred at many host locations. The recordings found that occurred before 2018 were incomplete with only the top 8 of tournaments recorded properly. This means that the tournament pools and top 64 games have been lost and unrecoverable. However, 2018 and 2019 has increased the number of records for most of the tournaments that are presented during the TWT. The number of tournaments in the 2-year period provide a good resource to track TWT results over this two-year period. 2018 was also the first-year tournament information was tracked on the official Bandai Namco website to highlight the TWT. Tournaments held in the USA were primarily observed for this research. The USA holds the most tournaments that are part of the TWT in relation to all the other host countries and were easily accessible. Evo Japan for 2019 and 2020 were also recorded as it was part of the EVO and TWT circuits. Being a part of the EVO series, many top players from the USA traveled and competed in these tournaments which enables data points for match records. A total of 18 tournaments were observed of the 61 tournaments that occurred over the world. A summary of Rounds, Games, and Sets are found in Table 17 for the tournaments. The last column shows the computation used the equation from Table 5 for double elimination tournaments which is 2n-1 to measure how many expected sets took place during each of the tournaments. All the tournaments followed a double elimination tournament format, and first to two game wins to determine the winner of a set. Table 18 provided the complete list of tournaments that were observed and part of the TWT (More fun for everyone!, n.d.).

Table 17. Summary Data of Observed Data

Tournament Name	Date	Country	Rounds	Games	Sets	# Registered	Max Sets
Final Round 2016	3/25/2016	USA	292	74	30	NA	NA
Evolution 2016	8/6/2016	USA	110	27	10	543	1,083
Evolution 2017	7/14/2017	USA	681	180	79	1,286	2,571
Final Round 2018	3/16/2018	USA	1142	291	126	297	593
NCR 2018	3/30/2018	USA	353	90	34	111	221
Texas Showdown 2018	5/4/2018	USA	211	55	20	150	299
Combo Breaker 2018	5/25/2018	USA	1048	272	121	488	975
CEO 2018	6/28/2018	USA	730	181	75	356	711
Evolution 2018	8/4/2018	USA	1081	287	125	1,547	3,093
Summer Jam 2018	8/31/2018	USA	645	163	68	228	455
SCR 2018	9/14/2018	USA	510	127	52	187	373
Evolution Japan 2019	2/16/2019	Japan	524	131	56	507	1,013
Northwest Majors 2019	4/26/2019	USA	506	125	51	161	321
Combo Breaker 2019	5/24/2019	USA	1248	318	131	627	1,253
CEO 2019	6/28/2019	USA	940	241	105	391	781
Evolution 2019	8/2/2019	USA	1110	298	130	1,899	3,797
Summer Jam 2019	8/30/2019	USA	666	169	68	263	525
Kumite In Texas 2019	9/13/2019	USA	112	27	10	200	399
Electric Cancel 2019	10/5/2019	USA	429	110	46	206	411
Dreamhack Atlanta 2019	11/15/2019	USA	790	201	85	335	669
Evolution Japan 2020	1/23/2020	Japan	580	149	59	963	1,925
TOTAL			13,708	3,516	1,481	10,238	21,468

Table 18. List of Tekken World Tour Tournaments with Date and Country of Origin

Name	Date	County
Final Round 2016*	3/25/2016	USA
Evolution 2016*	8/6/2016	USA
Evolution 2017*	7/14/2017	USA
Final Round 2018	3/16/2018	USA
Thaiger Uppercut	3/24/2018	Thailand
NCR 2018	3/30/2018	USA
Beast Arena Hong Kong	4/7/2018	China
Kuwait Battle Royale	4/12/2018	Kuwait
Korea Masters	4/28/2018	South Korea
Texas Showdown 2018	5/4/2018	USA
The Colosseum	5/12/2018	Italy
Battle Arena Melbourne	5/18/2018	Australia
Combo Breaker 2018	5/25/2018	USA
Fighting Games Challenge	6/9/2018	Poland
Taiwan Challenger	6/16/2018	Taiwan
ADFT	6/23/2018	Spain
CEO 2018	6/28/2018	USA
Abudget Cup	7/7/2018	Indonesia
FV X Sea Major	7/14/2018	Malaysia
VSFighting	7/20/2018	UK
Headstomper	7/28/2018	Denmark
Evolution 2018	8/4/2018	USA
Moscow Fighting Arena	8/11/2018	Russia
Tokyo Tekken Masters	8/18/2018	Japan
Celtic Throwdown	8/25/2018	Ireland
Summer Jam 2018	8/31/2018	USA
The Mixup	9/8/2018	France
SCR 2018	9/14/2018	USA
Rev Major	9/15/2018	Philippines
TXT	10/6/2018	Colombia
SEA Major	10/13/2018	Singapore
Berlin Tekken Clash	10/20/2018	Germany
Canada Cup 2018	10/26/2018	Canada
TWT Finals	12/1/2018	Netherlands
Evolution Japan 2019	2/16/2019	Japan

Table 18 continued

The Mixup	4/20/2019	France
Northwest Majors 2019	4/26/2019	USA
HeadStomper	5/3/2019	Sweden
RoxNRoll Korea	5/11/2019	South Korea
Battle Arena Melbourne	5/17/2019	Australia
Combo Breaker 2019	5/24/2019	USA
Electric Clash 2019	6/1/2019	Canada
Taipei Major	6/8/2019	Taiwan
Fighting Games Challenge	6/15/2019	Poland
Thaiger Uppercut	6/22/2019	Thailand
CEO 2019	6/28/2019	USA
Well Played Challenger	7/6/2019	Japan
VS Fighting	7/20/2019	UK
Evolution 2019	8/2/2019	USA
Collision	8/17/2019	Peru
FV Cup	8/24/2019	Malaysia
Summer Jam 2019	8/30/2019	USA
Cape Town Showdown	9/7/2019	South Africa
Kumite In Texas 2019	9/13/2019	USA
Clash of the Olympians	9/21/2019	Greece
REV Major	9/28/2019	Philippines
Electric Cancel 2019	10/5/2019	USA
Berlin Tekken Clash	10/19/2019	Germany
Tokyo Tekken Masters	10/26/2019	Japan
TXT	11/2/2019	Argentina
RoxNRoll Korea	11/8/2019	South Korea
Dreamhack Atlanta 2019	11/15/2019	USA
Tekken World Tour Finals 2019	12/7/2019	Thailand
Evolution Japan 2020	1/23/2020	Japan

Note. *not a part of the TWT

The games that were recorded used the game tracker report as described in section 3.1. While the games were being observed, a uniform partitioned overlay was applied to each of the two health bars that are displayed for player 1 and player 2. The health overlay was used to track the damage dealt and damage received at the end of each round. Each of the 20 partitions of the overlay reflected approximately 5% health to account for 100% starting health. Other factors that were tracked were which player was awarded first hit and the time remaining in the round.

The TWT was announced for the 2020 season, and the 2021 season, however during 2020 all face-to-face tournaments were canceled due to COVID-19. Evo Japan 2020 was the last face-to-face tournament before shutdown and was used as the last tournament for this study. Face-to-face tournaments were only considered for this study as online play tournaments can vary especially for a key metric that was being tracked that was side selection. Face-to-face play player is committed to either player 1 side (left) or player 2 side (right) for the starting positions and other in game metrics like their health bar, round win tracker, or special meters (character dependent). Online play for Tekken 7 allows the players to always select the side of their choosing for the player's respective screens.

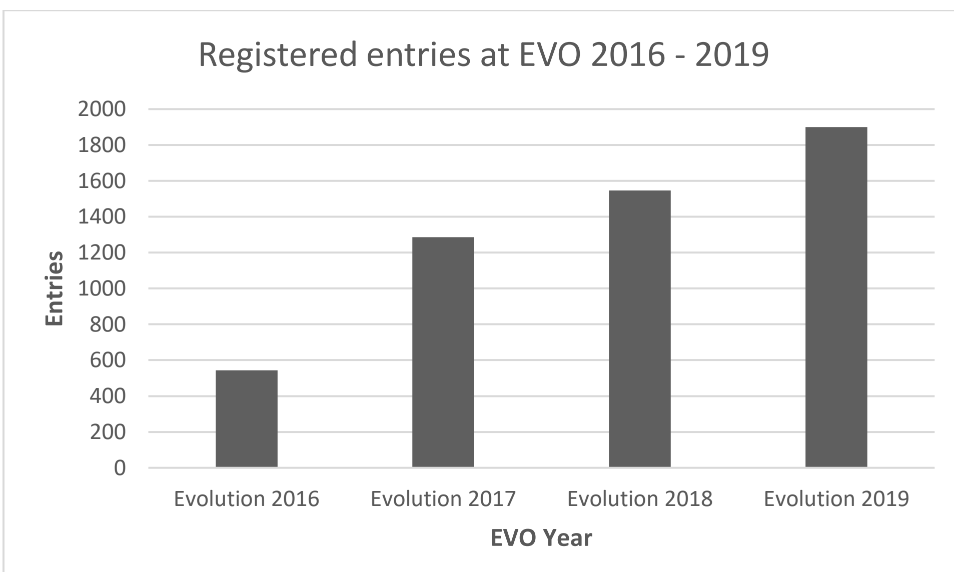

Figure 8. Registered Entries at EVO 2016 to 2019

4.3. List of Metrics Tracked from Tournaments

The full list of data that was collected during the game footage was described in Section 3.1. As matches were observed, the name of the players and the name of the characters for each respective side, along with all designed information were recorded. At the beginning of each round, the player that inflicted the first hit was recorded as a binary variable, as only one player can be awarded a first hit during every round. At the end of each round, the time remaining was recorded as well as any damage that remained using visual comparison to the uniform health bar overlay. All this data was organized in a parent spreadsheet that would be used for analysis in the following ways.

Characters selected were tracked to create a matchup chart. The matchup charts were to calculate the advantage (above 50%) or disadvantage (below 50%) character matchups that exists between two characters.

Player names were tracked as a category for further queries to calculate statistics for individual players and to later create a system of performance evaluation. The intended goal was to use player information for marketing purposes during a match. The next step was the creation of a predictive tool to aid spectators to the game.

Game rounds were tracked to keep an accurate count of how many rounds took place for a specific game. A player is determined the winner when three rounds are won in a first to three formats. A player has won decisively if three-to-zero rounds are won, and a close match occurred when three-to-two is the result. Every round that took place was assumed to be independent events for calculations. The magnitude in game decisions is measured by the amount of damage dealt and damage received. There was no in game feature to report the exact damage, hence an overlay was used with the health bars that was separated into 5% sections (20 sections) for each player. For instance, if health remaining is 12, that represents that health remaining was in between 55-60%. Special conditions are also tracked as they are perfects, time out, and double kos.

Perfects occur when either player takes no damage and takes the opposing player's health. These were tracked by damage remaining 20 and awarded a "p".

Time outs occurred after the round time limit of 60 seconds expires. The winner of the round was awarded to the player that has the most health in comparison to the other player. For tracking instances, if the two players had the same health bracket remaining, a decision was made to reduce the losing player down one bracket. For instance, if both players ended the round with a health remaining value of 5, the losing player would have their health bracket lowered to 4. If both players were in the 1 health remaining bracket, the winning player would have their health bracket increased to 2. The time remaining is recorded as "0" and a denotation of "t" is used.

A double knock out occurs when both players trade a damage blow while both being reduced to 0 health remaining. When a Double KO occurs, both players are awarded a round win and the next round continues. The denotation is "DKO" for this situation.

Table 19. Special Round Endings Statistics

	Perfects	Time Outs	DKO
Count of Rounds	717	82	15
Percentages of Rounds	5.23%	0.60%	0.11%

A game is a first-to-two (FT2) format, similar how rounds are tracked. Similarly, to the evaluation of round, a player who wins two-to-zero is explained as decisive win, while two-to-one is a close match. Health is tracked for magnitude of wins during each round. The Damage Dealt and Damage Remaining were used as a win predictor using equation 25.

A set is a combination of a first-to-two for games and first-to-three for rounds. Thus, in a set the minimum number of rounds that take place is six while the maximum is 15. Six is the minimum as a player can win decisively two times in a run, winning three-to-zero two times. Or the set was close by a player playing three games, with the round count of three-to-two, two-to-three, and three-to-two. Thus, the range for round is from six to 15.

First hit records the first player that dealt damage. If a trade occurred at the beginning of the match, the player that inflicted the most damage was awarded the first attack. If a trade of damage occurs and the damage appears to be equal, first hit was recorded when the next hit occurs. Chip damage was not recorded as a first hit, nor if a character uses an evasive ability at the cost of life. The magnitude of the first hit was not recorded as pausing to gain an accurate measurement of first hit damage would drastically increase the time that was used to observe data matches. If a game publisher were to code and publish the magnitude of first hit either during the round, or at the end of a round, this might provide a better glimpse of the impact of first hit, however, the data that is collected was to measure if there exists any type of impact of the round, game, and set.

4.4. Forecasting Matchup Results

The parent datasheet recorded 13,708 rounds and was utilized to identify patterns in the data to be used for predictive analysis. Three metrics were analyzed for comparison for an estimating set win % results. The primary objective that was used to measure accuracy of the forecast was comparing the player's overall set win % with the predictive indicators. Set win % was used as it is not the individual round or game that will matter in a tournament setting. The analysis that continued involved analyzing smaller metrics that make up the set win %. These metrics were game win %, round win %, and damage ratio.

Sets were composed of games, which were then composed of rounds, and rounds were composed of damage. These metrics are easily observable and provide insight whether they were good indicators. Pre analysis of these metrics began by querying the parent datasheet to create summarized data by categorical information, in this case, the player analysis data sheet.

4.5. Player Analysis datasheet

The data sheet used the player names (their identity) to organize the data. Upon review, there were two names that were used by two pairs of different players that ended up being thrown out due to naming conventions. The combined data remained in the player analysis datasheet since it was unclear of the total number of players that shared the same non-unique name. The player analysis held records of 895 players. The columns that were queried containing information related to the side, first hit, damage, rounds, games, and sets.

The players were sorted using number of sets recorded and set win %. The player with the most sets contained 81 sets and the number of sets that were recorded decreased down to the 48^{th} ranks to approximately 10 sets recorded. The number of recordings continues to taper from player 49 to 128^{th} from ~10 sets down to ~4 sets recorded. Further observations found that 595 players of the 895 players only have 1 set recorded (66.48% of observed samples). The situation occurs that for the top 16 players there are statistically significant figures, however the other data can be used. It was very common in Tekken 7 that a player will have few recorded data sets. Hence all the data for all the players will be used for predictive model. This was due to the nature of double eliminated tournaments, in which one-quarter of the tournament population begins to be eliminated starting in round two of the tournament brackets and continues until all players except one remain and is crowned the winner. The next subsets detail sets how to compute each metric to compare to the set %.

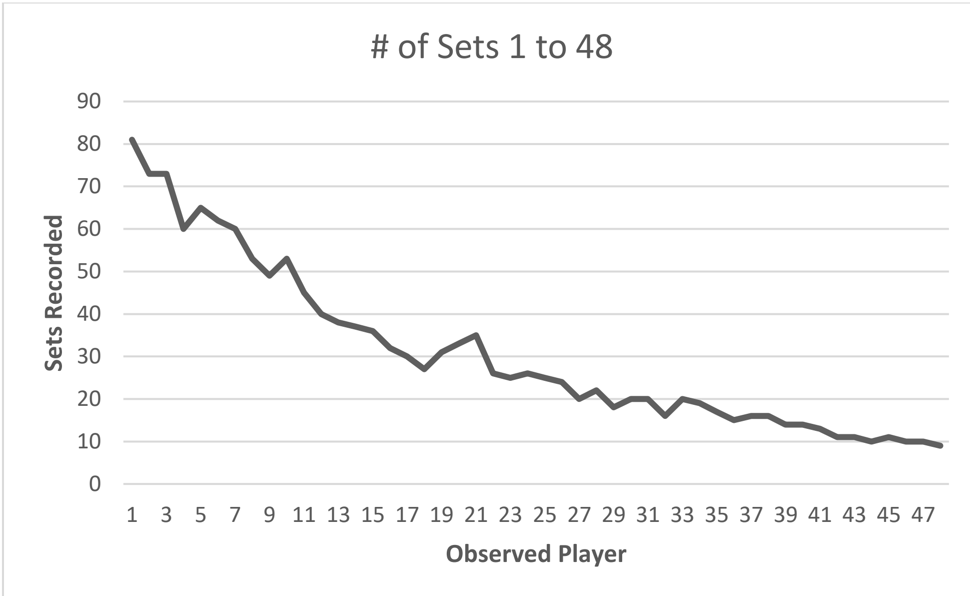

Figure 9. Number of Recorded Sets for the top 48 players

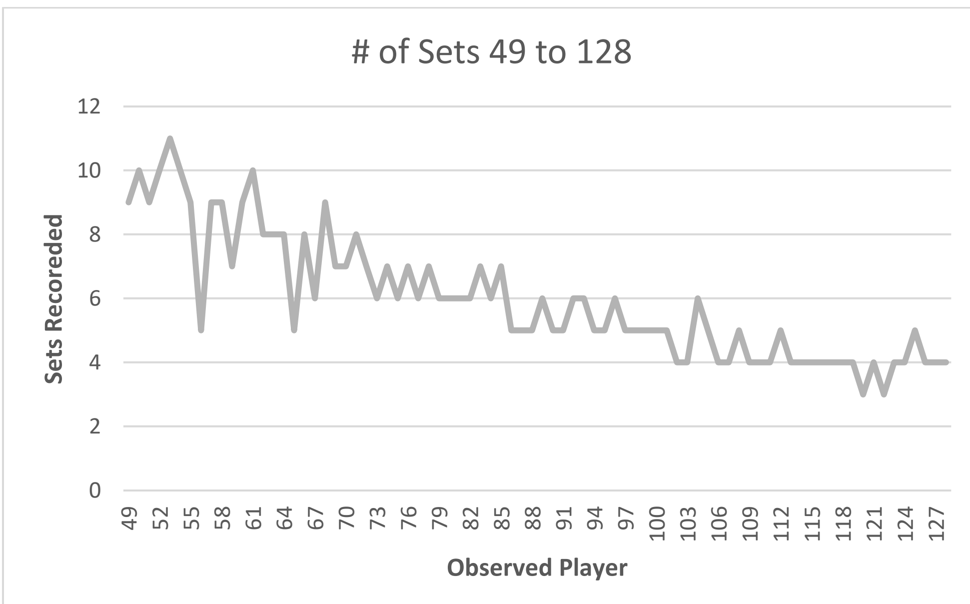

Figure 10. Number of Recorded sets for players ranked 49 to 128

4.6. Game Win %

Game win % (GW%) tracks the number of games won out of the games played. This was a standard metric players use to describe the result of a set. However, comparing the GW%

directly with the set win % has flaws. To attempt to overcome those flaws, further calculations are used to translate them into set form. A set was comprised of a first-to-two game wins. Section 3.8.1. described the probabilities that associated with all the different potential outcomes of games that can occur. Using the probabilities of a set, the win outcomes were used to have a direct comparison of game win % to set win %. The calculations are as follows.

Set Game Win %

$$SetGame\% = GW\%^2 + 2 * (GW\%^2 * (1 - GW\%)) \tag{19}$$

After incorporating the probabilities of outcomes of a set, a more direct comparison can be made for the set game % (SGW%) to the set% to be used as a forecasting metric. The SGW% is calculated for each player in the player analysis datasheet to be used to compute correlation, error metrics, and r-squared as described in Table 20.

4.7. Round Win %

The round win % (RW%), as with the game win % (GW%), modifications to the calculations must be made to be able to compare round wins more accurately % to Set win %. To convert the round win % to set win %; first is to convert RW% to GW%. The outcome of the game was completed as a first-to-three rounds. This was calculated as follows.

Round Win % Converted to Game Win %

$$GameRoundWin\% = RW\%^3 + 3 * \left(RW\%^3 * (1 - RW\%)\right) + \\ 6 * (RW\%^3 * (1 - RW\%)^2) \tag{20}$$

The calculations do not stop there, as the game round win% is now to be used as a comparison for game win %. The next step to be able to compare round win % to sets was to use the same calculations that was used to find set game %, but to use game round win % as the main variable instead of GW%.

Game Round Win % Converted to Set Round Win %

$$SetRoundWin\% = GRW\%^2 + 2 * (GRW\%^2 * (1 - GRW\%)) \qquad (21)$$

Converting the round win % to set round win% (SRW%), the direct comparison to set win % can be made. SRW% was calculated for each player in the Player Analysis datasheet to be used to compute correlation, error metrics, and r-squared as will be described in Table 20.

4.8. Damage Ratio

Calculating the damage ratio as described in equation 25. The parent datasheet containing the 13,708 rounds were used for the damage dealt and damage received. The information that was recorded was damage remaining for each respective character, and since there were 20 separated health bars to determine damage dealt. A new column was created that performed the calculations of 20-damage remaining. The new damage dealt column for both the sum of damage dealt and the sum of damage remaining.

Win Ratio from Damage Dealt and Damage Received per Round

$$\frac{Damage\ Dealt + Damage\ Remaining}{Rounds\ Played} = \frac{403{,}931 + 403{,}931}{13{,}708} = 58.9336 = x \qquad (22)$$

Observe that since there was a starting health (20) and health tracks down to zero (0), damage dealt, and damage remaining are equal. The reason that that simply do not equal 548,320 (13,708*2*20) is due to special instances when a Time out, or a double ko occurs. Now that is found, it is time to place it was the predictive equation as follows setting k=1, b=0 as these were smoothing constants:

Smoothing Constant for Damage Ratio

$$n^* = k * log\, x + b = log\, 58.9336 = 1.77 \qquad (23)$$

The next step is to set up an optimization function to find a best fit for the smoothing constants k and b. Creating an error forecasting charge to measure bias, mean squared error, mean square root error, and absolute error, to compare each players own winning round percentage and comparing to the forecast utilizing the damage ratio with n^*. Each player (895) was queried for the games that they played for both side 1 and side 2 to determine the individual player's damage ratio found by:

Damage Ratio

$$\frac{DamageDealt^{n^*}}{DamageDealt^{n^*} + DamageRemaining^{n^*}} = Damage\ Ratio \qquad (24)$$

After creating the individual player's damage ratios, the next step was to measure the error. The error category that we care most about is the mean squared error (MSE) as to minimize the sum of MSE is to minimize the magnitude of error. Solver optimization was used with the objective function to minimize MSE by changing the decision variables of b, and k the smoothing constants. The processing time took less than 1 minute to set a k=1.41827 and b=0.

$$n = k * \log x + b = 1.418827 * \log 58.9336 = 2.510858$$

Knowing the n-value of the sample size now gives the opportunity to have a starting measuring point. It also began the opportunity to use player's damage ratios as a predictive tool as they were matched up in a game. The predictive equation was calculated by using the individual player recorded damage dealt and damage remaining and using 2.51 is the n constant into equation 27, for all players.

4.9. Correlation, Error, and R-Squared

The previous calculations of prediction of set need to be measured for their accuracy. Set win % is the base for all the comparisons. It was imperative to find a metric that provides more data to feel confident in our predictions. In the player analysis datasheet, all the metrics detailed were used for further analysis for correlation, mean squared error, and r-squared Table 20. The summary table below displays these results.

Table 20. Correlation and Error Statistics of Recorded Rounds

Set%	Correlation	R-Squared	Mean Squared Error
Game%	0.912225	0.832155	21.299
Round%	0.865264	0.748682	41.412
Dmg%	0.864763	0.747815	46.26523
SetGameW%	0.928551	0.862206	17.50917
SetRoundW%	0.940305	0.884174	15.05961
Sd%	0.909855	0.827835	22.14877

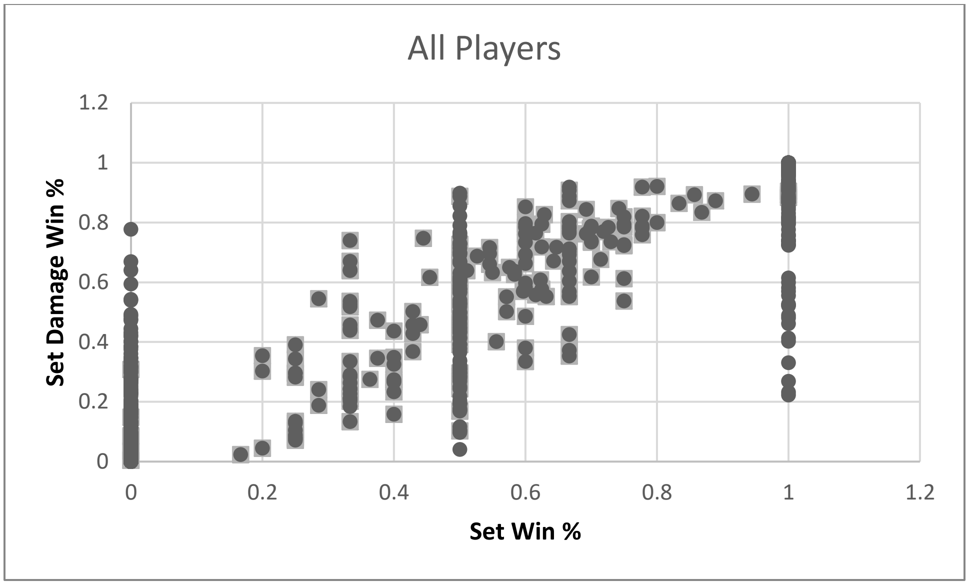

Figure 11. Set Damage Win % vs Set Win %

Table 20 demonstrates of these metrics, set round win % (SRW%) was the best indicator, followed by game win %, and set damage win %. Even though the damage ratio provided a n-value of 2.51, which is slightly higher than baseball's win% that ranges from 1.8 to 2.2. After calculation individual win rates for players, a matchup calculator of how to use these figures was investigated next.

Figure 12 demonstrates there was widespread since many players only have limited number of sets recorded. To eliminate these highly populated steps at 0, 50%, and 100%, a constraint can be added in which a minimum number of games was a requirement. Running r-squared analysis setting a minimum set yields the following results in Table 21. SRW% appears to be the strongest indicator expect in the case with 30 games in which SGW% was the highest.

Table 21. Summary of Min Sets and Metrics of Set Damage Win % / Set Round Win % / Set Game Win %

Min Sets	SDW%	SRW%	SGW%
ALL	0.827845	0.884174	0.862206
10	0.569191	0.755743	0.533066
30	0.442283	0.594032	0.612664
60	0.87864	0.943415	0.698874

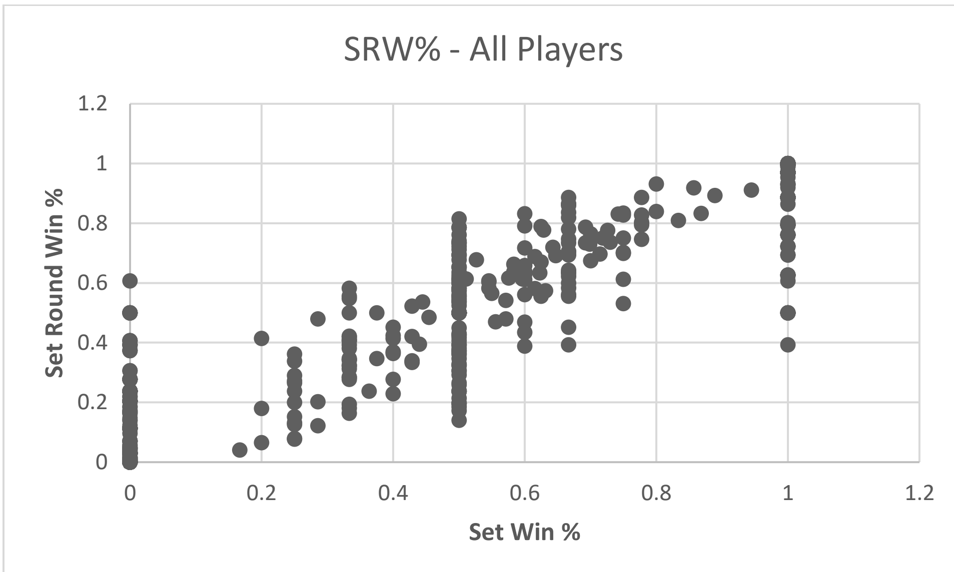

Figure 12. Set Round Win % of All Players

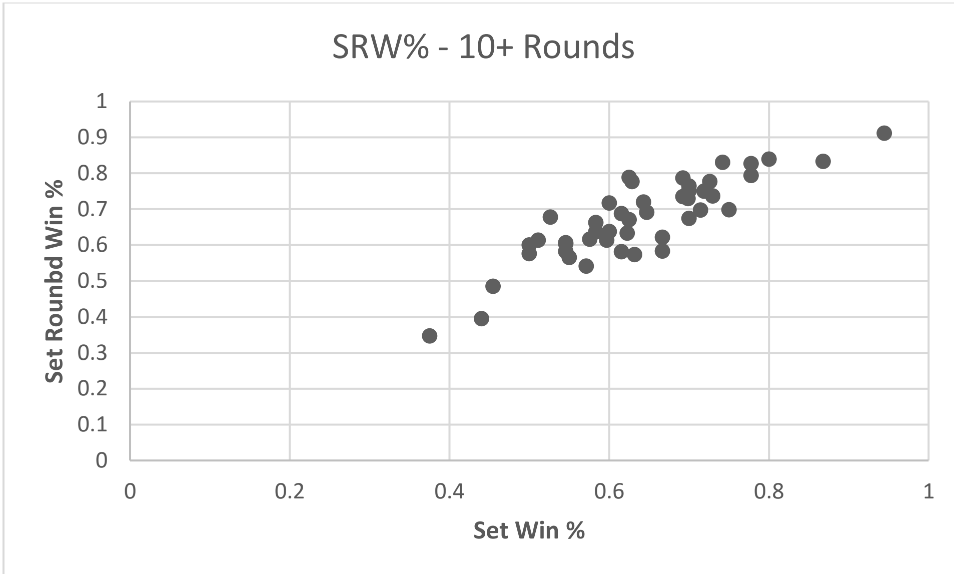

Figure 13. Set Round Win % Minimum of 10 Rounds Recorded

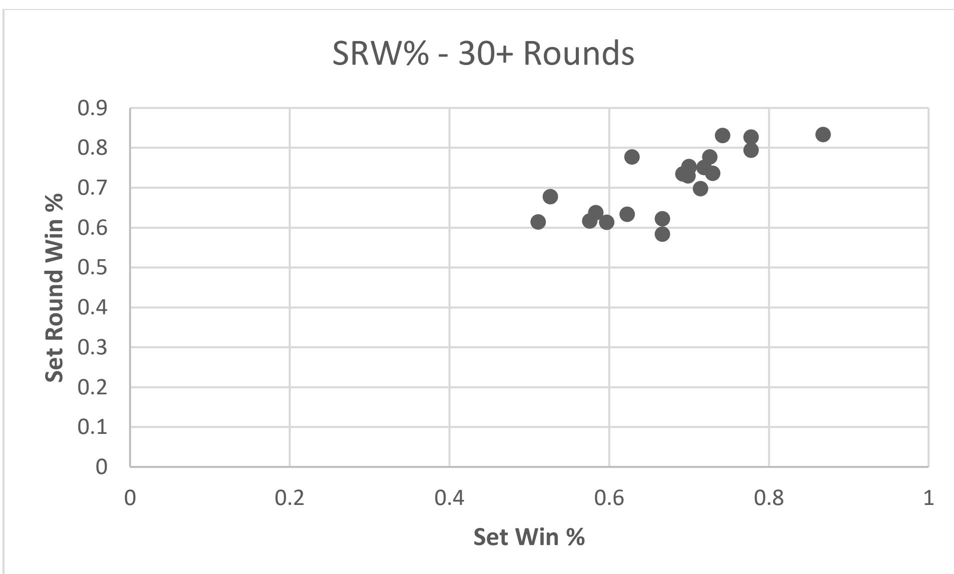

Figure 14. Set Round Win % Minimum of 30 Rounds Recorded

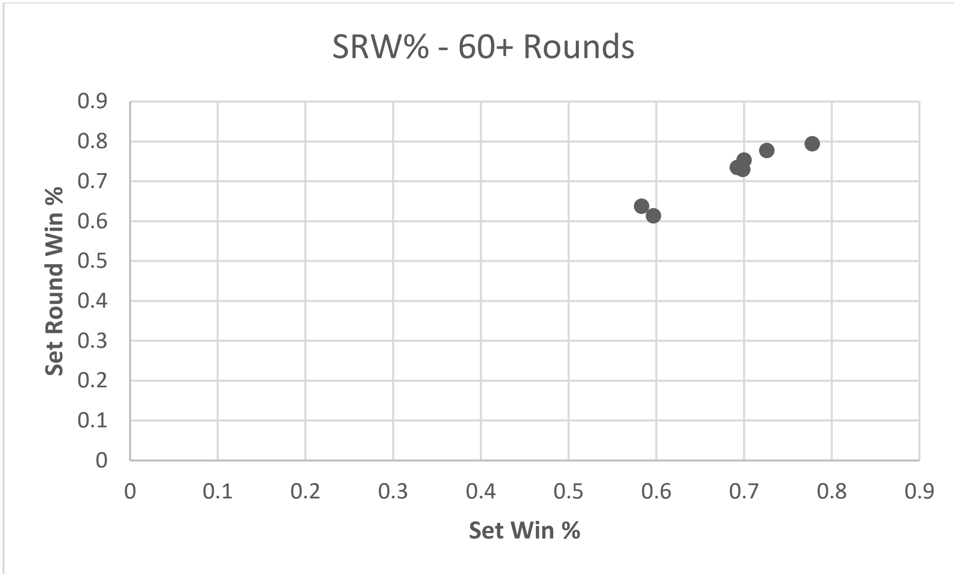

Figure 15. Set Round Win % Minimum of 60 Rounds Recorded

The best guess as to why there was a dip in accuracy around the 30 sets mark was that these players were more popular players. Popular players had more matches on stream at a higher frequency and showcase the players playing in the early rounds where they have a large advantage against their opponent. As there are more advantage games of these players recorded, there is the possibility of their data being skewed. Without a complete picture of a player for the start of a tournament to the end of their run, and a large enough sample size, there was insufficient recordings to explain this outcome without speculation.

4.10. First Hit as a Predictor

The first hit was tracked and used to see how the effects of scoring a first hit impacted game performance. The first hit as a round predictor saw no correlation between round %. Recorded first hit is compared to round win % as a first hit occurred during each round. Further analysis, an outcome matrix can be shown that scoring a first hit on average did increase a player's chances of winning the round by approximately 10%. While on average this was true, this was highly dependent on the player. Of the players that have at least 30 rounds recorded (152 players), the highest first hit average was 66.7% with a first hit win % of 45.2%, however this player was categorized as the 109th player. The outcome matrix that is found below uses all data and calculates the conditional probabilities for scoring a first hit, and the conversion to a win or lose. Further breakdown analyzes non-pool play, top 64 to top 8. The last table show top 8 only. This was done to compare if there are any trends that stand out between the different player ranks. Findings were similar to Sirlin's findings in his article 'Game Balance and Fantasy Strike' (2019), in which he analyzed character advantage matchup between similar rank groupings between lower ranks and higher ranks had little variance to each other for these global statistics.

Table 22. First Hit Win % All Rounds

Global	Y	N	
W	29.84%	20.22%	50.06%
L	20.16%	29.78%	49.94%
	50.00%	50.00%	

Table 23. First Hit Win % Top 64 Player Rounds

Top 64	Y	N	
W	29.53%	20.52%	50.05%
L	20.47%	29.48%	49.95%
	50.00%	50.00%	

Table 24. First Hit Win % Top 8 Player Rounds

Top 8	Y	N	
W	29.22%	20.85%	50.06%
L	20.78%	29.15%	49.94%
	50.00%	50.00%	

Table 25. Player with Highest Win % with First Hit each Round

109^{th}	Y	N	
W	45.2%	14.3%	59.5%
L	21.5%	19.0%	40.5%
	66.7%	33.3%	

The results of first hit data and the effect it has on forecasting a round win is very small. However, it was observed that scoring a first hit does have an effect by increasing the average player's chance of winning the round.

4.11. Matchup Calculations

The application of the prediction element needs explained how to use it in real matches. An issue with using rounds as a predictor in comparing two players was the number of stepwise calculations that occur. Previously observed was Figure 12, there are many players that fall in the 0%, 50%, and 100% brackets. To make a more continuous variable was to use the set damage win% (SDW%). SDW% measures the magnitude of a win and a loss, and an advantage for comparing two players together was even if a player falls within a 0% set round win % (RW%), the comparison was not always calculated that the player will lose. SDW% was able to have some magnitude data, so even with the finding in the previous section, when there is not enough data, the confidence of using magnitude data should be used since theoretically it was always possible for one player to win no matter what.

To use the forecasting metrics for direct comparison of players, a reference matrix was created to look up information in the player analysis datasheet for a player and their opponent. The summed percentage would not equal to 100% but was equal to 200% since player 1 was equal to 100% and player 2 was equal to 100%. To measure the expected win rates in the matchup, there needs to be an equation to set the values sum to be equal to 1. To find the adjusted win percentages, first identify the player's win probability and loss probability. Next divide that sums by 2 (as there are two players) to have the newly adjusted win probabilities for both players. The win probability comes from the damage ratio.

Probability of Player 1 winning vs Player 2

$$P_{1,W} = \frac{WP_1 + (1 - WP_2)}{2} \tag{25}$$

$$P_{2,W} = \frac{WP_2 + (1 - WP_1)}{2}$$

$$P_{1,W} = \frac{0.629 + (1 - 0.55)}{2} = 53.9\% = Player1RoundWin\%$$

$$P_{2,W} = \frac{0.55 + (1 - 0.629)}{2} = 46.1\% = Player2RoundWin\%$$

Utilizing the damage ratio provides the information of the expected win rate for each individual round. Through this analysis, an assumption that is made is that each round was independent of one another. The next step was to convert from rounds into games, and finally into sets.

The round win percentage was taken and looking at the original structure of tournaments and games, it was known that Tekken games are played in a first-to-three round format. A round probability chart can be formed and the round win% can be used to complete the respective win/loss for the players.

Table 26. Format to Calculate Game Win % from Round Win %

Game Probability(Round)				
R1	R2	R3	R4	R5
P1RW%	P1RW%	P1RW%		
P1RW%	P2RW%	P1RW%	P1RW%	
P1RW%	P1RW%	P2RW%	P1RW%	
P2RW%	P1RW%	P1RW%	P1RW%	
P1RW%	P2RW%	P2RW%	P1RW%	P1RW%
P1RW%	P2RW%	P1RW%	P2RW%	P1RW%
P1RW%	P1RW%	P2RW%	P2RW%	P1RW%
P2RW%	P2RW%	P1RW%	P1RW%	P1RW%
P2RW%	P1RW%	P2RW%	P1RW%	P1RW%
P2RW%	P1RW%	P1RW%	P2RW%	P1RW%
P1RW%	P1RW%	P2RW%	P2RW%	P2RW%
P1RW%	P2RW%	P1RW%	P2RW%	P2RW%
P1RW%	P2RW%	P2RW%	P1RW%	P2RW%
P2RW%	P2RW%	P1RW%	P1RW%	P2RW%
P2RW%	P1RW%	P2RW%	P1RW%	P2RW%
P2RW%	P1RW%	P1RW%	P2RW%	P2RW%
P1RW%	P2RW%	P2RW%	P2RW%	
P2RW%	P2RW%	P1RW%	P2RW%	
P2RW%	P1RW%	P2RW%	P2RW%	
P2RW%	P2RW%	P2RW%		

While this information can be used for display purposes, the main pieces of information that were needed for computation were the win rates for Player 1. The computation for game win percentage is as follows for each of the winning combinations. Player 2 Round Win% (P2RW%) can be expressed as 1-P1WR%, they are equivalent.

Player 1 Game Win % using Round Win %

$$Player\ 1\ Game\ Win\ \% = P1RW\%^3 + (P1RW\%^3 * P2RW\%^1) * 3 + (P1RW\%^3 * P2RW\%^2) * 6 \qquad (26)$$

$$Player\ 2\ Game\ Win\ \% = 1 - (Player\ 1\ Game\ Win\ \%)$$

After finding the probabilities for the game win % for each of the players, it is time to perform the last step by computing the set win percentage. Following the same steps as was performed for the game win percentage, with the change that sets are determined in a first-to-two set result format.

Table 27. Format to Calculate Expected Set Win % from Game Win %

Set Probability		
G1	G2	G3
P1GW%	P1GW%	
P2GW%	P1GW%	P1GW%
P1GW%	P2GW%	P1GW%
P1GW%	P2GW%	P2GW%
P2GW%	P1GW%	P2GW%
P2GW%	P2GW%	

Same as before, this is good display information, but to simplify out the computations as follows:

Player 1 Set Win % using Round Win %

$$Player\ 1\ Set\ Win\ \% = P1GW\%^2 + (P1GW\%^2 * P2GW\%^1) * 2 \qquad (27)$$

$$Player\ 2\ Set\ Win\ \% = 1 - (Player\ 2\ Set\ Win\ \%)$$

4.12. Game Preparation Decision Making

Improving in a fighting game can be difficult. Training and study a fighting game, or any game, takes determination and will to learn the mechanics, tactics, and overall strategies to win. To train for tournament play, player need to learn all of that, within the constraints of the tournament. The extensive form game for fighting game tournaments have been detailed out section 3.2. The next topic explored were the decisions that occurred before a game as related to side selection, and character select I/II/III.

Up to this point, the predictive analysis and rating have been focused on the player during play for game I/II/III. No in-depth review was public for making out of game decisions except a few top players that appear to be involved with the activity as observed in tournaments. Side selection has always been a contention to the point of including rules for a player to request the right to play rock-paper-scissors (RPS) to determine which players sits on their preferred side. The observed data of the top 128 players there was a breakdown of characters selected as found in Table 28. These characters are not all used in a single tournament, as player may decide to switch characters for a variety of reasons, however, the data suggested there was an opportunity to improve and the steps to make the decision to practice and change characters.

Table 28. Number of Characters that are used by Tournament Players

# of Characters Used	# Players Count
1	55
2	40
3	19
4	6
5	2
6+	6

4.13. Side Selection Subgame

The beginning of every set has the players determine what side each player was seated on, player 1 side (left) or on player 2 side (right). The official Evo rules state that players can ask to play rock-paper-scissor (RPS) for a player to decide what side that they are seated at (Evo Tournament Rules, 2020). The question that is explored was, if there exists a real in game benefit of sitting on side 1 or side 2?

The analysis began by observing the player analysis and measure the global round win % for side 1 and side 2. Even though there were 13,708 rounds recorded, the total recorded wins were equal to 13,724. This was due to the possibility of double kos exist in which both players were awarded a win for the round. Side 1 recorded a total of 7,506 while side 2 recorded 6,218 round wins. Taken as a percentage, it was observed that the win % for side 1 is 54.7% and side 2 was 45.3%. There does appear that there is a slight advantage on side 1 for the side selection.

To see if the side advantage is global, next only top players only were analyzed to identify if the global advantage remains true, or if all players favor side 1 in respect for win %. The top players that were measured, only 38 players had at least 30 rounds for both side 1 and side 2. Of these top players, a side advantaged was observed for side 1 round win % at 61.8% and round win % at side 2 52.9%. Side 1 round wins at 4,416 out of 7,145 rounds and side 2 round wins at 3,200 out of 6,049 rounds. The top player averages are above 50%, and being top players, the round win % should be above 50% to be on top, and there appears to be an advantage for side 1. This advantage is not true for all players. Of the 38 players, 25 favored side 1 with 5% or more, 10 players are side neutral, and 3 players performed better on side 2 with a difference over 5%.

From the side selection analysis, most of the players' favor side 1. If the sample size in the top can be extrapolated to the rest of the ranks as they were in section 4.10. with respect to first hit for pool play, top 64, and top 8, based on these results will follow a first mover advantage game.

The first mover advantage game to gain this advantage should be played as follows. When a match is being called, and placement is determined, attempt to be the first player to take a seat on side 1. The opponent reserves the right to challenge the selection by asking for rock-paper-scissor (RPS) but may not always ask for it based on behavioral patterns. If a player sees a player seated on side 1, always call for RPS for the side.

4.14. Character Select I/II/III Subgames

The character select subgames I/II/III are similar in nature, however the calculations for decision making were vastly different to maximize character advantage. The next sections describe the logic that was used at each stage of the subgames to maximize the character advantage. Identify equilibrium points in each of these subgames, first was to analyze the data with respect to characters to identify tier lists and matchup charts.

Matchup charts and tier lists are explained in section 1.15. Not only are modern tier lists using expert subjective base decision making, modern tier lists also use summative scores for decision making. Moreover, the tier lists were used primarily as a discussion piece comparing other personal anecdotal evidence. As noted in section 2.9. Sirlin created his own tier list for Fantasy Strike and discussed briefly about the data. During this research, the recorded data were categorized in a similar matter, but the breakdown and interpretation goes further than the summative data that was the generally accepted. Furthermore, the matchup charts are used to exploit the decision making that was used in a tournament setting to identify what character to practice with, and potentially what characters to practice against. In short, the decision before a game even begins was important and using matchup chart data assists with the decision.

4.15. Character Datasheet

The first step of the character datasheet was the filter all the characters that were used in tournament recorded matches. At the time of recording, the last character that was released in the time frame was Leroy Smith, and at least one set of each character was recorded. Not all character combinations have been observed. Only 1,217 matchups were overserved to have at least one set recorded out of the 2,304 possible matchups. The meaning is that only 52.8% of matchups were observed to add in decision making and analysis. 634 matchups (27.5%) are observed when 13 rounds are the minimum, and 212 (9.2%) matchups have 30 or more rounds recorded. The lack of data required additional assumptions to explain the steps and interpretation of the charts, however some analysis was too impractical. The analysis that was not completed due to lack of data is discussed in section 6.3.

4.16. Matchup Charts

Creating a matchup charts, sufficient data needs to be gathered to understand each matchup to produce a sufficient confidence interval of the gathered data. The confidence interval of everyone may vary based on their own preference, but a target of 68%, 90%, and 95% are the most used. To have each of these confidence intervals are satisfied at, 300, 500, and 600 data points for each matchup are needed to feel confident of the amount of data that is gathered. Table 17 shows that there should be sufficient data during a tournament, only to find that a small proportion of data is recorded for analysis. It was impossible to capture this data that is lost to time, thus further assumption must be made. It is assumed that players that participate in these tournaments to some degree take matchup charts into consideration based on their own individual conclusions of matchups. This assumption falls in line with early stated assumption that a player will select a character that maximizes their chances of winning. Players could gather their own data, or collectively gather data using a network of players, in ladder matches or private matches to exploit specific character matchups to conclude what character maximizes their probability of winning.

The matchup chart described above was interpreted with the advantage (above 50%) or disadvantage (below 50%) that a character has against one another. These figures have in combination the stage, side, and player rating. Further analysis using principal component analysis (PCA) was attempted to measure the impact of each of these factors, however too much data was missing for PCA to be used even under probabilistic PCA. In addition to the individual matchup charts, are the character usage. The chart can then be simplified by removing any character that is not in use, categorized as dominated characters, to assist in further analysis of the selected tournament characters if there exists further dominated and thus to be removed. By removing and isolated these options, as a player was looking to participate in a tournament the data provides more accuracy to support a decision of who to play and who to practice against to increase the player's chances of winning. Having a precise metric to practice against was expected to be a better utilization of their time, compared to practicing a matchup or scenarios that they may not experience in tournament.

4.17. Decision of What Character to Select

The next section described how to selection a character during the subgames character selection I/II/III. There was a robust system to maximize the probability of winning in the upcoming game through the subgame character selection. The two strategies that were analyzed came from the categorization of player decisions for character selection to be either rule of cool (RC) or top tier (TT). RC is simply the selection of the character based on a subjective decision for each individual player, without the full evaluation of a character's advantages and disadvantages. The TT player will select using a more objective approach toward matchup chart analysis. The TT will analyze tool sets a character possess and the win percentages that were attained with that character. The next section will break down how the RC and TT players make decisions and how each will perform.

The addition of categorizing the players requires, three more terms are introduced to assist with the decision theory for character selection I/II/III. No matter if the player is RC or TT, they will select a lead character (LC). The LC was chosen under the different constraints of the player. A RC player selects a LC based on subjective qualities, while a TT player will select the LC using the tier charts. Both players know the extensive form game of a tournament and have their own character pools (CP) for both soft counter characters (SC) and hard counters characters (HC). A HC is a used in which knowing the opponent's character, a character is selected that has the highest advantage versus that selection. SC looks to maximize their advantage of the known opponent's character selection, and takes into consideration how their opponent will counter pick the SC. As it was easier to measure the effects of HC selection compared to SC, character selection III will be observed first, and backward induction was used to assist selecting best response character selection to subgame character selection II and character selection I.

4.18. Character Selection III

Character selection III began with the situation in which both players have won a game each, and now one player must select the same character that won the last game, while their opponent now has the decision to change their character. For this example, assume that player 2 was making the selection and assume that player 2 can play all characters as they are presented in the matchup chart. Player 2 will then utilize the matchup chart to identify what character their opponent has selected, and in response, player 2 will select the character that will provide the most advantage knowing their opponent character. As there are no more options for their opponent to select in character selection IV, selecting the hard counter character (HC) is the best response during character selection III. Relaxing the possibility that player 2 can play all characters, player 2 will look up their character pool and from their character pool and select the HC from their character pool.

Player 1 Hard Counter Selection

$$Player\ 1\ Hard\ Counter = MAX(P(P1HC|P2SC)) \tag{28}$$

4.19. Character Selection II

Continuing backward induction, character selection II occurs after game 1. Following the example in the previous section, the scenario for character selection II player 1 has the option of selecting a character in response to losing to player 2 after game 1. Player 1 now faces the decision and must anticipate their opponent's potential future response. Meaning if player 1 uses a hard counter character (HC) without the consideration of player 2's response, it is possible that player 1 maximizes their character advantage for game 2 but may face an even tougher HC during game 3 when it would be player 2's turn for character selection. To gain an advantage for game 2 and minimize the response of Player 2's, the SC is the best option. To select a SC, reference the character matchup chart and select the pair of characters by player 1's selection and multiplying the HC of that pick. This is the combined probability of the player 1's selection and player 2's response. For sake of argument, the same counter pick method can be used for player 1, and if the player 2 counter pick (P2CP) is known, that can also eliminate player 2's response.

Player 1 Soft Counter Selection

$$Player\ 1\ Soft\ Counter = MAX(P(P1SC|P2LC) * MIN\big(P(P2HC|P1SC)\big)) \tag{29}$$

4.20. Character Selection I

The last decision for the backward induction analysis was selecting a character for character selection I. Character selection I was unique since there can be two instances that occur. The first scenario is a double-blind pick is called by one of the players. A double-blind pick was an option that either player may request before a character was selected. Any player may ask for a judge to use the ruling for a double-blind pick in which one of the player's tells the judge in secret what character the player wants to select, while the other player selects a character. After the player selected a character, the judge then announces the character that must be picked. The second scenario is the lead character counter pick is to eliminate the situation in which one or both players appear to be waiting for their opponent to select a character, and then immediately is able to make a character selection to counter pick that character which begins the counter pick algorithm during character selection I.

4.20.1. Character Selection I – Double Blind Pick

A double-blind pick was described in the previous section as the first scenario. To make a decision that needs to be made for both player’s is to define and use their lead character (LC). The LC for each player is the character that each individual feels would provide them with the greatest chance of winning, and the greatest compliment in their character pool. There are two main ideas that were explored in section 4.17 that there are rule of cool (RC) players and top tier (TT) players.

The RC player will select a LC based on some subjective parameters. The compliment characters will use the subjective LC as the starting point for the remainder of the character pool. The TT player has analyzed the matchup chart and will select their characters based on maximizing the character advantage of possible character selection for their opponents. Traditionally, summative score is used for identifying the top character. However, knowing the established player responses, a LC should be based on expected value that was derived from matchup chart analysis. Since the opponent character is unknown, due to a double-blind pick, all the opposing LC are under uniform probability of being selected. It has already been computed what the responses will be using the soft counters and hard counter strategy for character selection for each individual LC selected. The strategies used were constructed from the matchup chart analysis by selecting a LC that will maximize the expected win %.

4.20.2. Character Selection I – Lead Character Counter Pick

The double-blind pick was put in place to stop the practice of lead character (LC) counter pick. The situation occurs when one player selects a character, and their opponent observed the selection and in response directly selects a character to have an advantage matchup. The question this research is interested in is what type of counter pick should occur? The same expected value that was used for the double-blind pick was used for the evaluation. Gaining the advantage of observing the opposing LC guarantees for game 1 there was an advantage. The calculation of the expected value changes from a uniform distribution of character listing, and instead sets the observed character chance of a matchup to 100% for game 1. The probability follows the tournament structure for expected character advantage used in equation 33. The equation was constructed and found below to calculate the expected value during subgame character selection I.

Player 1 Expected Character Advantage Counter Picks

$$Player\ 1 ExpectedCharacterAdvantage = (P1LC|P2LC * P2SC|P1LC) +$$

$$(P1LC|P2LC * P2SC|P1LC * P1HC|P2SC) +$$

$$(P1LC|P2LC * P1SC|P2LC * P2HC|P1SC) \tag{30}$$

4.21. Character Matchup Chart Reduction

The rule of cool (RC) player will rely on their lead character (LC). After side selection has taken place, the RC already knows the main character that they will select regardless of who their opponent's character selection is. Section 4.20.1 explored the expected win without the consideration of the opponent. The logic laid out in the creation of Table 3 is the same logic used for matchup up chart reduction. The same disadvantage exists from Table 3 as they are with the recorded Tekken 7 data, however, the conclusion does not mean that the RC player is acting irrationally. One of the assumptions that were made was a player will make decisions to maximize their probability of winning. However, the RC player acts rationally as the time was a larger constraint for the RC players and want to speed up the selection process without matchup chart analysis. When time becomes a factor the RC players were not only making the selection of a single character based on subjective criteria, but the decision of not picking up an additional character has the potential of not having the time commitment needed to learn a new character. The RC has a risk of selecting a character that underperforms because of inherent toolsets when compared to other character options due to lack of analysis of available options.

The situation occurs then when new players are joining that have a limited amount of time available to them that the player wants to minimize this risk and planning of a character based on the top tier (TT) decision making. This approach looks to reduce the risk of dedicating time in a character that has more disadvantages than advantages. Outside of a new player that was minimizing risk with the goal of winning in mind rather than the exploration of a character based on how cool they look or the interesting backstory, established players look at individual matchups to gain specific advantages. Simplified, tier lists are analyzed to determine a lead character, and matchup charts are used for players that want to add characters to their respective character pool. With that in mind, character selection I comes in terms of who to practice and create a character pool to minimize the risk and increase the player’s probability of winning. Subjective picking a character is not defined in this study, however the character pool can be defined using matchup charts to increase their probability of winning. TT decision making will be emphasized in the remainder of this research.

A new tier list using the soft counter entries was then created using equation 33 for all matchup possibilities. The data table is too large to display in this research, however a TT player would have to learn and play 17 characters for maximum character advantages. The new tier list uses the game theory to reduce and calculate the opponent's best responses and to limit the maximum counter play that would given to the opponents. This was vastly different when compared to the traditional tier lists as were described in Table 3 using a summative scale for evaluation. The new tier lists enable to identify not only one character but enables the player to create a top tier character pool that compliments other character selected based on the possible responses of future opponents.

The 13,708 rounds that were recorded were used to create a matchup chart. There are 48 characters to select thus a 48x48 matrix was created to matchup chart. Once the matrix was created, only 1,217 of the possible matchups were observed of the possible 2,304; approximately 52.8% of the matchup chart was created. This count of matchups takes into consideration if there is at least one occurrence of the matchup. If a constraint is applied further to only observe matchups that have at least 30 rounds, the observed matchups go from 1,217 down to 212 (9.2%). While only a small selection of matchups was recorded during tournament play, this information needs to be interpreted.

30 rounds were selected to begin to have a minimum sample size of 30 to show statistical results. The new matrix was reduced from a 48x48 matrix down to a 35x35 matrix. Further elimination in which there are only 1 matchup found for a character are reduced further from 35x35 matrix down to a 25x25. This reduction took place as the matchup chart is to help identify

what character to select for during the counter pick phase, if there was only one character found, it the only observation is either above or below 50%-win percentage, and no other comparisons. With the most reduced matrix being 25x25, the next step is to observe any characters that are dominated by another. To identify dominated characters, first identity the best character response for player 1 given that player 2 will select each of the remaining 25 characters. The characters for player 1 that have a recorded highest value remained as they were calculated to be the best responses. Next, begin with a character for player 1 that does not have a best response to any player 2's character selection, and beginning comparing to other characters that player 1 was able to select. To describe strict dominance, each individual matchup for a specific character was compared with the same matchup set to different available characters. When the situation occurs that all variables in the set are greater than the initial character set, strictly dominance was meet. If strictly dominance was observed, the dominant character is recorded and removed for the available selection of characters. This process was conducted for 12 characters, of these 12 strictly dominance was found in 5 of these characters. Removing the strictly dominated characters creates the new matchup chart to be a 20x20. The 20x20 matrix has 400 matchups that are compared, and 163 matchups recorded (~40%), that have at least 30 rounds. Results of the reduce matrix were outline in Table 29.

Table 29. Reduced Matchup Chart with Hard Counters Highlighted

	Akuma	Alisa	Asuka	Bryan	Claudio	Devil Jin	Dragunov	Eddy	Geese	Hwoarang	Jack7	Jin	Josie	Kazumi	King	Law	Paul	Shaheen	Steve	Xiaoyu
Akuma				0.278	0.513							0.647								
Alisa									0.353		0.513					0.481				0.385
Asuka							0.375		0.289	0.362	0.444		0.581		0.543	0.400	0.525		0.452	
Bryan	0.722						0.432	0.441	0.343		0.455			0.432	0.257					
Claudio	0.487							0.186		0.500										0.419
Devil Jin							0.569	0.407	0.396		0.481			0.533		0.647		0.533	0.633	0.610
Dragunov			0.625	0.568		0.431		0.600	0.447		0.505	0.620		0.674	0.578		0.481		0.568	
Eddy				0.559	0.814	0.593	0.400				0.318	0.415		0.510	0.605		0.442			
Geese		0.647	0.711	0.657		0.604	0.553			0.600	0.469	0.391		0.281		0.535		0.533	0.452	0.514
Hwoarang			0.638		0.500				0.400		0.455	0.424					0.472		0.372	
Jack7		0.487	0.556	0.545		0.519	0.495	0.682	0.531	0.545	0.500	0.634	0.491	0.579	0.579	0.596	0.417	0.538	0.586	0.603
Jin	0.353						0.380	0.585	0.609	0.576	0.366			0.390	0.613			0.459	0.563	
Josie			0.419								0.509				0.769		0.394		0.543	
Kazumi				0.568		0.467	0.326	0.490	0.719		0.421	0.610			0.444			0.645	0.463	
King			0.457	0.743			0.422	0.395			0.421	0.387	0.231	0.556						0.333
Law		0.519	0.600			0.353			0.465		0.404								0.500	0.409
Paul			0.475				0.519	0.558		0.528	0.583		0.606							
Shaheen						0.467			0.467		0.462	0.541		0.355						
Steve			0.548			0.367	0.432		0.548	0.628	0.414	0.438	0.457	0.537		0.500				
Xiaoyu		0.615			0.581	0.390			0.486		0.397				0.667	0.591				

The data collect was incomplete, and thus must be interpreted as such. First, as there were missing matchups in tournament, it was assumed that these characters in the selection process have been deemed lesser by the tournament participants. This was assumption is supported by top 8 and higher bracket matches not identifying these characters. This can be for several reasons, all of which will not be explored further in this study. Second, was to eliminate more characters that have insufficient data or comparative matchups in the matchup chart. Thirdly, using game theory to eliminate strictly dominated characters. Lastly, a count is made of the known matchups that are above 50%, and if at least half of the reduced matchup charts were recorded. Appling these filters provides the information of which characters were the top tier characters and have the highest chances of being considered a lead character in tournament play. The case of Tekken 7 yields three characters 1) Jack7 with 13 of 18 known matchups being above 50%, 2) Geese with 9 of 13 known matchups being above 50%, and 3) Dragunov with 8 of 11 known matchups being above 50%.

After the top 3 characters were identified under this decision-making process, next was to estimate the win probability distribution of each of these characters. Until this point, it was assumed for deciding for player 1 to follow the top tier decision making. Now it was assumed player 2 knows the options available and will act according. The different perspective was analyzed to identify the best response for player 2 with respect to the future response of player 1. This process will attempt to identify potential equilibrium points and create character counter pick chains.

4.22. Creating a Character Pool for Maximizing Matchup Advantage

There are several different approaches a player can take when deciding on their character pool to use during a tournament. Adding characters to a player’s character pool has the benefit of taking advantage of character matchups and a focused approach on how to deal with different threats. A player that is making informed decision with the use of either the global statistics and/or their own personal statistics. The data that has been gathered globally from the tournaments is used and some information is incomplete. The matchup knowledge is null as there have been no video records of the matches that took place. In the case where information is null, it was assumed that a player will default to their lead character (LC) for character selection. When information is present, a player will change characters to have at least a character advantage defined by a round win percentage that is above 50%.

Character pool approach 1 is to find a combination of characters that is the theoretical best. This approach assumes that a player has enough time to devote and learn each character to at least the global average win percentages. The two scenarios that occur are 1) blind pick is called during character selection 1, and 2) a player was allowed to counter pick starting before game 1 begins.

Scenario 1 is when players request blind picks. Blind picks occur when one player makes a hidden character selection and the opponent selects there character publicly, and the hidden selection is shown and selected. The LC for a player to select under uniform distribution was different for each possible LC. Incomplete data gives the situation a variety of other constraints that can be applied to the scenario. A constraint was added for comparisons were only characters that have minimum data recorded to update the matchup chart, with sufficient data of at least half character were considered. Eight characters fit this criterion and of those characters, the win rate ranges from 40.8% to 50.9%. The character King has a 50.9%-win rate under blind pick conditions and instead of needing to learn 40 characters for this strategy, the total characters in the character pool decrease to 17.

It is better for a player to begin counter picking before game 1, and that was what occurs in scenario 2. The figure that was calculated is the theoretic max expected win% that can be obtained through countering picking as described in equation 33. The theoretic win% with perfect countering using both soft counters and hard counters comes to 61.9% (with removing 50% matchup due to insufficient data for the 7 characters), and 60.2%. The total amount of characters that will need to be learned to achieve this win percentage is 40. Comparing the 50.9% in blind pick vs the 60.2% that is gained as an advantage of beginning counter picking before game 1.

A conclusion for this could be that a blind pick selection should be used whenever possible, however, a known player versus an unknown player potential gives the advantage to the unknown player that can take known public information and unknown players make a blind counter pick to the unknown player to still gain the advantage. If there are two known players and a blind counter pick was selected, there is a separate subgame of known potential characters and can calculate the proper response, however this subgame is not explored in this paper.

Any number of constraints can be added to this framework, including amazing if a player only has time to learn and practice 2, 3, or 4 characters and measuring each phase of the margin benefit of adding characters. However, the nature of incomplete data of not having every matchup calculate with a minimum of statistically significant number, it is up to the discretion of a player to determine what constraints to add to the system. Even with missing data, previously three characters were identified as the top tier characters. If a player were to select each of those three characters, while their opponent uses the SCHC tactics, 3) Dragunov win rate goes from 56.3% to 33.1%, 2) Geese 54.6% to 26.9%, and 1) Jack7 60.3% to 43.7%. Analysis should be conducted for each of the characters, however since of the lack of data, only these top 3 characters were used for comparing the marginal benefit chart to include additional characters to a player's character pool.

4.23. Stage Selection

When it comes to stage selection, both players and characters must adapt to the environment. In Tekken 7, stages have a variety of features such as walls, balcony breaks, wall breaks, floor breaks, or have no walls. Using the ballpark factor adapted from baseball (equation 34), that measured if a stadium is advantages to offense or defense, the damage ratio is used to find the stage factor for each of the stages. This was computed by taking all the damage that has been recorded on each respective stage and comparing it with the other stages in relative terms. The stage factor was used as a multiplier for character selection proposes. It can also identify when a player was having less than or greater than the average recorded damage value. The stage factor helps identify the stage selection process, and how gaps in knowledge are identified for players and characters.

When a player has the option of selecting a stage the goal to maximize their probability of winning a round. Having the knowledge of the character matchup, a stage may be selected to help benefit the selecting player while minimizing the effect of aiding the opposing player.

On a more individual level, certain players may not have practiced the specific features that occur on these stages and can be exploited for their lack of knowledge about stage factor. The situation will occur when a closer examination of player specific knowledge may be required, but the overall character knowledge was easier to use as a deciding factor simply based on the information was more readily available for characters, oppose to individual player records. This was due to the fact many players can play the same character that is used to create the damage scores when determining the stage factor when compared to individual players.

If a player does not follow the counter picking methodology, or if the player is locked in with a single lead character, the stage select can play a role into maximizing the difference in their character's benefit versus the benefit and hindrance in their opponents. For instance, it is thought that King performs best on the stage Forgotten Realm, and is supported by this stage being picking two time more than any other stage, but the win percentage on the stage is just about 50%, while the stage Hammerhead has 50 rounds played on the stage and has a win percentage of over 60% suggesting that King players should not auto lock into Forgotten Realm, as there were plenty of other character's that can benefit from the floor break mechanic but do struggle on longer stages that require wall carry which make Hammerhead a possibly better option.

Stage Damage Multiplier

$$100 * \left[\frac{\left(\frac{homeDamage\ Dealt + home\ Damage\ Received}{home\ Rounds}\right)}{\left(\frac{awayDamage\ Dealt + away\ Damage\ Received}{away\ Rounds}\right)}\right] \tag{31}$$

Table 30. Stage Damage Adjustment

Stage Name	Stage Factor
Abandoned Temple	100.6389
Arctic Snowfall	99.8850
Arena	99.1146
Brimstone & Fire	100.4507
Devil's Pit	101.2628
Dragon's Nest	100.5754
Duomo di Sirio	101.2338
Forgotten Realm	98.8032
G Corp Helipad	99.7522
Geometric Plane	99.9112
Hammerhead	100.0676
Howard Estate	100.1038
Infinite Azure	100.4588
Jungle Outpost	100.1808
Kinder Gym	99.8663
Last Day on Earth	98.1972
Mishima Building	98.7033
Mishima Dojo	99.8059
Precipice of Fate	100.4658
Souq	99.0466
Twilight Conflict	100.4964
Violet Systems	99.8951

Up until now, only characters have been used in the analysis of stages, and that was where it will stop. Player knowledge does play a role, but it is assumed that a tournament player will have knowledge or to understand how to use the tools of a character to use the environment to their advantage or to minimize the impact of their opponent’s character. Lower ranked player would have the tools available, but the lack of knowledge would be recorded in the globally tracked damage dealt and damage ratio metrics.

The strategy of selecting a single character and playing the stage counter pick game can be explored. The same issue with the counter picking, there are many incomplete or insufficient data in the stage chart to have an accurate ready. Instead of running through the results, the same method that was previously used now incorporates the information of a matchup and the stage that was selected. For instance, if Geese and Dragunov were playing on Forgotten Realm, Geese's damage is at 84.4% while Dragunov's damage is a 106.2%. This difference was unclear until principal component analysis can be conducted to have a deeper understanding of how stage factor, character factor, player factor, and opponent factor all part take in a game outcome.

4.24. Summary

Chapter 4 explored the data that had been gathered and the steps and interpretation of that data. The main topics included how to predict a player matchup utilizing round win % and introduction of the damage ratio as a prediction tool. Discussion continued with the decision making that is required to maximize side and character advantage by creating character pools. The end of the chapter discussed how stage can play a role, but due to lack of comparable data, no further analysis could be down to show significance between the decision of selecting a stage or selecting a character during character selection I/II/III subgames. Chapter 5 will summarize finding that have been covered in Chapter 4.

CHAPTER 5. SUMMARY AND OUTCOMES

5.1. Overview

This chapter summarized what has been explored in Chapter 4. The summary decisions and results that focus on game theory decision making. Chapter 4 focused on the fine details of the data; Chapter 5 examines the results and what decisions the players should make that utilize the information that was gathered.

5.2. Data

There was a glaring issue with how matches are recorded. As shown in Table 17, only 1,481 sets were recorded, compared to the calculated 21,468 that should be available from the tournaments that were observed for during this study. That made only 6.9% of data expected versus what was available to be used for evaluation research. Further analysis of Table 17 concludes that less than 10% of expected sets to recorded are preserved for historical purposes. The conclusion to draw from the recorded is to find alternative to identify an improved tracking mechanism

5.3. How to Forecast the outcome of a match

Making a prediction between two players have shown results. Methodology was created to find what indicators are correlated and have statistical significance to predicting a set outcome. It was found that round win % was the best indicator in predicting the outcome of a set, followed by game win % and finally the damage ratio. The unexpected was that the damage ratio ranked 3rd in the study when it was a expected the damage ratio would be at least 2nd. However, the correlation of 86.4% found in Table 20 demonstrated that the metric can still be used.

Because a high correlation was found for the damage ratio should be explored further as a potential indicator for future research.

While there was insufficient data to perform principal component analysis to create a metric that combines the factors of player, character selection, opponent selection, and stage, the idea should not be abandoned. Framework for each of these metrics were identified, and while some metrics like stage factor are sufficient, the other factors are lacking and require further data gathering.

5.4. How is performance rated

Performance was rated using the Set Round Win % found in equation 24 for individual players. The player performance was attached to each of the players, and when comparisons were used between two players, a breakdown was available based on the expected values presented in equation 28. Method used is in position for further research to evaluate the accuracy of the prediction method.

Further evaluation of individual metrics was expected, however insufficient data would not support any findings for performance. The main methodology was to utilized principal component analysis to identify the impact of each of the metrics. When player matches are recorded, a combination of these metrics can be broken down to factor what the player’s overall abilities are to potentially have a more accurate measurement on performance.

5.5. How to maximize decision making during Tekken 7 Tournament Play

Forecasting a matchup focused on the aspects of the extensive form for Game 1/2/3. To maximize a player’s decisions also are impacted before the game starts. Evaluation of all subgames was the primary focus to draw conclusions. The subgames analyzed were side selection and character selection I/II/III.

The side selection subgame found that there was an advantage to the player that is on side 1. The subgame explanation concluded that there is a first mover advantage that has the possibility of taking advantage of behavior aspects in which a player will not call for rock-paper-scissor to be called, and the winner of rock-paper-scissor to choose their preferred side. The situation then showed if a player is already seated on side 1, call for rock-paper-scissor to attempt to win side 1. The reasons for the side 1 advantage are unclear, as in game decisions were not analyzed. There is speculation that specific actions are easier to perform from the player 1 side when compared to player 2 side. Regardless, the data analyzed in this study concludes there is a side 1 advantage and take advantage of it when possible.

The next groups of subgames were related to character selection I/II/III. Backward induction was used in the evaluation of character selection based on the possible character picks of the opponent. Equation XYZ was used for the computation of different mix of characters in a character pool for the evaluation of responses to maximize character advantage. A brute force algorithm was used to compute the variety of different combinations that can take place during the character selection subgames. Different style of players was defined and taken into consideration which were the rule of cool (RC) players, and the top tier players, in the formation of the player respective character pool. The intention of this section was to identify what character and character pool will maximize the character advantage. A conclusion to draw is that a player should only ever use one character.

Matchup charts were the main way to identify how to make the decisions of character selection, and how to establish a lead character, soft counters, and hard counters for player's individual character pool. It was noted that the direct application of learning how to play all character seemed impractical, there were a measurable marginal character advantage that follows a logarithmic curve of marginal gains by adding characters to the character pool.

The character selection subgame I was analyzed under two different scenarios, 1) double-blind pick, 2) counter picking, this can impact what character to add to the pool. Under double-blind pick, it was observed that having Asuka as a lead character and learning 16 additional characters will provide the highest character advantage at 59.2%. The counter pick scenario would create a max character pool of 40 character with an expected character advantage of 60.2%. Thus, utilizing the double-blind pick reduces the characters needed to be practices by more than half, with the marginal loss of ~1% of character advantage. When all possible character specialization was compared, Jack7 has the highest character specialist advantage of a measurement of 43.7%. To plot all the possible combinations of character pools was excessive to perform for this paper. There were 8.4*10^12 combinations for double blind pick, and 2.8*10^14 combinations for counter picking in subgame I. The equations and logic laid out can be used for the examination of all possible combinations, but that will be representative of individual players that want to calculate their own result. The conclusion was to set up an optimization model in that case to limit the number of characters in the pool and compare the results always to player 2 that does not have the added constraints for the best result.

5.6. Summary

This chapter summarized the findings from chapter 4. The process and results of the data, forecasting a matchup, player ranking, and maximizing subgame decisions were discussed. Next chapter 6 will cover future research and discussion of the topic.

CHAPTER 6. DISCUSSION AND FUTURE RESEARCH

Chapter 6 is reserved for future research and discussion points that were related Tekken 7 tournaments. Due to lack of data during this research study left some on questions asked. There are also possible concerns of the analysis of the data and how it can relate to the ever-changing landscape of the Tekken 7 and how the same steps can be used for a variety of fighting games or other games all together. The follow is the discussion of those points and highlights potential future research.

6.1. Discussion

A potential issue with interpretation of this data was there is no consideration of updates or balance adjustments. As it has been previously stated, a lack of data for character matchups evaluations leaves the determination of conclusions difficult. The research laid out the approaches and methods to observe this type of data and organize it for evaluation. Balance updates were not partitioned in the data that could play a role in the character selection decisions to take place. Instead of creating a new character pool for each balance update, a fixed character pool was made and to measure the long-term evolution of characters through all balance updates.

Another point was the evaluation of the stage factor. Equation 34 and Table 30 demonstrated how to implement the ballpark factor for each of the characters. The section then did not incorporate the findings into decision making, other than if not changing character, to select the stage that provides the maximum difference between the selecting character and the opponent. Stage selection for this research is assumed to be a lesser factor than changing character from the character pool, because the principal component analysis was incomplete.

Further analysis could identify the perceived factor that the stage has on the characters and this metric would be added in the evaluation of character selection I/II/III.

6.2. Further Research

An observation of research is the direct relationship of player vs player without any other factors. The history of digital fighting games has had rivalries develop over time. Some of these rivalries continued to show that one player always won, even though the other player would consistently beat other better players. A future study would be interesting to investigate what makes a good rival story and what type of decision making occurs in which one player is "in the head" of their rival. The emotional response to these situation or what game impact situations occur that impact the decision making of their opponent.

Outside of digital fighting games, the tournament structures need further research. Detailed in Table 5 detail a set of different tournament formats. Due to the varying sizes and volume of in which tournaments occur, this provides an opportunity for further evaluation of tournament formats. While the current digital fighting game tournament landscape has long been double elimination, does not discount the possibility of experimenting with other tournament formats.

An operations research focuses research can be constructed for a variety of constraints that will help in the selection process of tournament style based on a variety of constraints, such as time, space, number of player stations, security, and financial constraints. The fighting game community, while has improved over the years, still have a lot to learn about how to run a tournament including Evo 2019 in which a group of players were unable to pass security in time and were disqualified due to waiting in line for reported amounts of two hours. For some of these players to travel to another country to participate in a tournament only to face a loss due to managerial lack of knowledge certainly should be inexcusable.

6.2.1. Principal Component Analysis

Principal component analysis (PCA) observes the data to identify primary characteristics by reducing the categorizing. The data collected measured character matchups, stage selection, and side selection, the three important factors when it comes to increase likelihood of winning a round in Set play in Tekken 7 tournament setting. Unfortunately, the research data was unable to utilize this reduction technique. There is some method of PCA that like the probability PCA to estimate unavailable data, but typically those methods have a threshold of have 15% or less data that is being estimated. The biggest grouping that was missing data was nearly 50% for matchup chart analysis. The problem was further exacerbated when using players as the key elements, as very few players have many matches recorded.

The PCA should use players and track the individual percentages to be on the same scale with respect to side, stage, characters against. While players will have a character pool significantly less than the entire available characters, the distribution of opponent characters should be a fair representation of the character distribution in tournaments. There are two approaches to increase the matchup data.

The first approach is for tournament organizers to track this information when reporting the results of a set. Providing the information of games and rounds provides more information to be used for analysis. While it is more difficult to attain the damage ratios for further testing, there would at least be more data to analyze rather than reporting a set win and sometime not even the game wins and losses. The second approach is to be provided a dataset from Bandai Namco. It can be assumed that Bandai Namco tracks pieces of this information. During online play, Tekken

7 displays the global stats for different character matchups, along with personal matchup win %. This information if shared can be used as an estimate for tournament play, and ultimately PCA evaluation. Online play does have some differences that may impact the offline stats, however, over the other option of having no data at all.

A significance point of view, Bandai Namco can use the interpretation of this information to attempt to draw attention to their game. A continual question that new players ask is who they should play? This research encourages to play anyone but take in mind to make a character pool that is complimentary to that character. In other words, instead of only character trailers, some information can be presented in a manner that compliments a player's lead character and provides insight into matchups that are possible perceived as unfavorable. Giving options to unfavorable matchups has the potential for more engagement in the community by creating a feedback loop through balance update changes.

6.3. Summary

This chapter has concluded the documentation of this study be looking towards future opportunities of research. Digital fighting game research is having growing interest from the social point of view, and the purpose of this research is to quantify some of the decisions that are required related to side selection, character selection, and stage selection. This research for Tekken 7 provided a framework for future analysis of other digital fighting games. All digital fighting games can utilize game theory approaches to maximizes chances of winning based on measurable results. It is this hope and continuation in this space to not only gain an understanding for competitive play, but to highlight the uniqueness of digital fighting game tournaments that can host over one thousand participants at a time provides opportunity for future research.

REFERENCES

Abdi, H., & Williams, L. J. (2010). *Principal component analysis.* Wiley interdiscplinary reviews: computational statisctics.

Albert, J. (2010). Sabermetrics: The past, the present, and the future. *Mathematics and sports*, 3-14.

Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior, 8*(1), 6-19.

Azar, O. H., & Bar-Eli, M. (2011). 2011. *Applied Economics, 43*(25), 3591-3601.

Becker, B., & Huselid, M. (1992). The incentive effects of tournament compensation systems. *Administrative Sceince Quarterly*, pp. 336-350.

Ben-Naim, E., Rdener, S., & Vazquez, F. (2007). Scaling in tournaments. *EPL (Europhysics Letters), 77*(3), 30005.

Boice, J. (2018, June 13). *How Our 2018 World Cup Predictions Work*. Retrieved November 2, 2020, from FiveThirtyEight: https://fivethirtyeight.com/features/how-our-2018-world-cup-predictions-work/

Bouton, C. L. (1901). Nim, a agame with a complete mathematical theory. *The Annals of Mathematics, 3*(1/4), 35-39.

Chu, C. C., Chang, T., & Chu, J. (2016). Opposite Hand Advantage and the Overrepresentation of Left Handed Players in Major League Baseball. *Academia Economic Papers, 44*(2), 171.

Connelly, B., Tihanyi, L., Crook, T. R., & Gangloff, K. A. (2014). Tournament theory: Thirty years of contests and competitions. *Journal of Management, 40*(1), 16-47.

Deck, C., & Kimbrough, E. O. (2015). Single-and double-elimination all-pay tournaments. *Journal of Economic Behavior & Organization, 116*, 416-429.

Duke, R. (2018, January 24). *Understanding Standings, Part I: Tournament Structure (The Basics)*. Retrieved November 2, 2020, from Channel Fireball: https://strategy.channelfireball.com/all-strategy/mtg/channelmagic-articles/understanding-standings-part-i-tournament-structure-the-basics/#:~:text=In%20a%20Swiss%2Dstyle%20tournament,players%20who%20would%20like%20to.&text=Players%20earn%203%20match%20po

ELO, A. E. (1978). The rating of chessplayers, past and present. *Arco Pub.*

Empowering esports communities. (n.d.). Retrieved November 2, 2020, from smash.gg: https://smash.gg/

Evo Tournament Rules. (2020). Retrieved November 2, 2020, from Evo Shoryuken: Evolution Rules https://evo.shoryuken.com/tournament-rules/

Fusco, A. (2012, September 24). *Why Royce Gracie is the most Influential MMA Fighter of All Time*. Retrieved November 2, 2020, from Bleacher Report: https://bleacherreport.com/articles/1345700-why-royce-gracie-is-the-most-influential-mma-fighter-of-all-time

Gibbons, R. (1992). *Game Theory for Applied Economists.* New Jersey, Princeton: Princeton University Press.

Glickman, M. E. (1995). The glicko system. *Boston University, 16*, 16-17.

Groothuis, P. A., & Hill, J. R. (2018). Exit Discrimination in Major League Baseball: 1990-2004. *Southern Economic Association, 75*(2), 574-590.

Guan, Y., & Dy, J. (2009). Sparse probabilistic principal component analysis. *In Artifical Intelligence and Statistics*, 185-192.

Guha, R. (2003). *Open rating systems.* Almaden: IBM Research.

Harper, T. (2010). The art of war: Fighting games, performativity, and social game play. *Ohio University*.

Harper, T. (2013). The culture of digital fighting games: Performance and practice. *Routledge*.

Henz, M., Müller, T., & Thiel, S. (2004). Global constraints for round robin tournament scheduling. *European Journal of Operational Research, 153*(1), 92-101.

History of UFC. (n.d.). Retrieved November 2, 2020, from UFC: https://www.ufc.com/history-ufc#:~:text=The%20goal%20was%20to%20find,sumo%20and%20other%20combat%20sports

James, B. (1981). *Baseball Abstract.* Lawrence: self-published.

Júnior, P. S., Gonçalves, M. A., Laender, A. H., Salles, T., & Figueiredo, D. (2012). Time-aware ranking in sport social networks. *Journal of Information and Data Management, 3*(3), 195.

Ketonen, M. (2016). *Designing a 2D fighting game.* Kajaanin Ammattikorkeakoulu University of Applied Sciences.

Koller, D., & Pfeffer, A. (1997). Representations and solutions for game-theoretic problems. *Artifical intelligrence, 94*(1-2), 167-215.

Larkey, P., Kadane, J. B., Austin, R., & Azmir, S. (1997). Skill in games. *Management Science*, 596-609.

Leone, M. (2014, February 3). *Street Fighter II an Oral History*. Retrieved November 2, 2020, from Polygon: https://www.polygon.com/a/street-fighter-2-oral-history

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game.* New York and London: Norton.

Liao, H. (2006). A brief historical review of Olympic urbanization. *The International Journal of the History of Sport, 23*(7), 1232-1252.

Mann, C. (2009). Gladiators in the Greek east: a case study in romanization. *The international Journal of the History of Sport, 26*(2), 272-297.

Maskin, E. (2011). Commentary: Nash equilibrium and mechanism design. *Games and Economic Behavior, 71*(1), 9-11.

McHale, I. G., Scarf, P. A., & Folker, D. E. (2012). On the development of a soccer player performance rating system for the English Premier League. *Interfaces, 42*(4), 339-351.

Meehan, E. (2020, April 9). *The True History of Brazilian Jiu Jitsu*. Retrieved November 2, 2020, from BJJ Success: https://www.bjjsuccess.com/history-of-brazilian-jiu-jitsu

More fun for everyone! (n.d.). Retrieved November 2, 2020, from BANDAI NAMCO Entertainment America: https://www.bandainamcoent.com/tekken-world-tour

Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the national academy of sciences, 36*(1), 48-49.

Pearson, K. (1901). One lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2*(11), 559-572.

Rabinovich, Z., Naroditskiy, V., Gerding, E. H., & Jennings, N. R. (2013). Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. *Artificial Intelligence, 195*, 106-139.

Raymond, M., Pontier, D., Duguor, A., & Moller, A. (1996). Frequency-Dependent Maintenance of Left Handedness in Humans. *Biological Sciences, 263*, 1627-1633.

Rothman, S. (2014). A New Formula to Predict a Team's Winning Percentage. *Baseball Research Journal*.

Santos, R. (2019). A Forecasting Model of Success at the Euro Tournaments: The Role of Team's Performance at Qualifying Games. *Journal of Applied Business and Economics, 21*(8).

Shepanik, A. (2015). Graph labelings and tournament scheduling. *Doctoral dissertation, University of Minnesota*.

Sirlin, D. (2014, September 16). *Balancing Multiplayer Games, Part 1: Definitions*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/balancing-multiplayer-games-part-1-definitions

Sirlin, D. (2014, September 16). *Balancing Multiplayer Games, Part 4: Intuition*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/balancing-multiplayer-games-part-4-intuition

Sirlin, D. (2014, September 17). *Solvability*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/solvability

Sirlin, D. (2015, March 26). *Balancing Multiplayer Games, Part 3: Fairness*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/balancing-multiplayer-games-part-3-fairness

Sirlin, D. (2015, May 4). *Game Balance and Yomi*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/game-balance-and-yomi

Sirlin, D. (2017, May 7). *Balancing Multiplayer Games, Part 2: Viable Options*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/articles/balancing-multiplayer-games-part-2-viable-options

Sirlin, D. (2019, November 7). *Game Balance and Fantasy Strike*. Retrieved November 2, 2020, from Sirlin.net: https://www.sirlin.net/posts/game-balance-and-fantasy-strike

Snowden, J. (2018, November 12). *UFC 1, 25 Years Later: The Story Behind the Event that Started an Industry*. Retrieved July 14, 2021, from Bleacher Report: https://bleacherreport.com/articles/2804552-ufc-1-25-years-later-the-story-behind-the-event-that-started-an-industry

Team, B. (2020, June 9). *Swiss Tournament System - What Is It? How Does It Work?* Retrieved November 2, 2020, from Bitspawn: https://bitspawn.gg/swiss-tournament-system-what-is-it-how-does-it-work/

The Impact of Street Fighter 2. (2021, Februrary 8). Retrieved May 4, 2021, from Game Rant: https://gamerant.com/street-fighter-2-fighting-games/

Tipping, M., & Bishop, C. (1999). Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61*(3), 611-622.

Trick, M. A. (2002). Integer and constraint programming appraoches for round-robin tournament scheduling. *International Conference on the Practice and Theory of Automated Timetabling* (pp. 63-77). Berlin: Springer.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior, 2nd rev*.

Wang, M. (2019). *The Values of a Fighter*. Retrieved November 2, 2020, from Student Association for Applied Statistics Berkeley: https://saas.berkeley.edu/rp/the-values-of-a-fighter

Yang, Y. F., Feng, Q. Y., Sun, Y. L., & Dai, Y. F. (2009). Dishonest behaviors in online rating systems: cyber competition, attack models, and attack generator. *Journal of Computer Science and Technology, 24*(5), 855-867.

VITA

Kurtis Homan

Education

Doctor of Philosophy (Ph.D.) in Supply Chain Technology 05/2014 – 08/2021

Purdue University

Master of Business administration 08/2012 – 05/2014

Purdue University

Bachelor of Science in Management and Quantitative Methods 08/2004 – 05/2009

Illinois State University

Professional Experience

Purdue University West Lafayette, IN 05/2014 – 05/2020

Instructor

- Designed course curricula in manufacturing and supply chain systems including case studies, activity-based learning activities and exams for undergraduates
- Taught 16 sections of manufacturing and supply chain systems to undergraduate students in the School of Technology, and one section of global logistics and transportation for senior undergraduate students
- Mentored teaching assistants on supporting class instruction, homework grading, and handling student needs
- Partnering with Rolls-Royce, created a student final project using specific designed case studies targeted to analyze Rolls-Royce's decision regarding production planning
- Transitioned to online learning effectively after COVID pandemic with implementation of recorded lectures including tutorials and both synchronous and asynchronous virtual classrooms

Research Assistant 05/2015 – 08/2016

- Assisted Dean of Technology to increase alumni engagement by creating a demographics database and created surveys to increase School of Technology Alumni engagement
- Benchmarked other university programs to update the masters, Ph.D., and faculty handbooks to be approved by School of Technology senior leadership

National Science Foundation 2015 Grant

- Constructed active learning activities to demonstrate activities related to logistics and supply chain management for middle school and high school students

Wabash National Corporation, Lafayette, IN

Annual Manufacturing Summit Guest Speaker 06/2018

- Conducted a 4-hour workshop with activity-based learning activities for the Senior Leadership team related to process improvement focusing on identifying high volatility processes that impact the system throughput using statistical analysis
- Guided discussions on the Theory of Constraints to be implemented on the manufacturing floor, including implementation of forecasting methodologies for workforce planning of manufacturing scheduling

Indiana Next Generation Manufacturing Competitiveness Center, West Lafayette, IN 05/2014 – 08/2016

Consultant - Bosma Enterprise, Indianapolis, IN

- Created and built a computer assisted design of the manufacturing plant to meet the needs of a workforce that is 50% visually impaired

Consultant - Company Confidential

- Initiated a Time in Motion study to measure assembly times and identify bottleneck processes which resulted in an increase throughput of facility by 37% without additional financial investment

AGCO Corporation, Batavia, IL

Supply Chain Development Intern 05/2013 – 08/2013

- Implemented a process improvement initiative at a newly acquired inbounding center (APIC) focused on workflow process that realized 10% cost reduction on labor by appropriate utilization of temporary staff
- Developed an inventory management program that forecasts inbound inventory and assist with staffing decisions at APIC
- Trained the warehouse manager at APIC to train associates how to use forecasting information to utilize floor space
- Conducted business analysis for the strategy team for the selection of a product information management system to be used organization-wide to implement an e-commerce strategy

Independent Consultant

LynxIT Solution, Westmont, IL 06/2011 – 09/2011

- Created and implemented central SQL database to track physician schedules, help desk information, and sales status
- Transformed data directories for physicians, practices, and hospital clients allowing for ease of finding and updating information to a centralized location
- Designed automated administrative reports to track sales and employee effectiveness using SQL

Touro Infirmary Hospital, New Orleans, LA 05/2010 – 08/2010

- Lead the Medical Staff Office through re-design of the medical staff credentialing and privileging processes
- Nominated new Medical Staff Leadership to support and facilitate the change plan
- Streamlined multiple database processes to reduce errors related to physician information
- Orchestrated change in the bylaw procedures which included education for medical staff members and employees

Rush Copley Medical Center, Aurora, IL

Data Analyst 05/2007 – 05/2010

- Designed and supported special projects through a variety of business functions that include, Philanthropy, Billing, Medical Records, Physician Services, and Physician Offices focused on process improvement initiatives
- Collaborated with Director of Revenue Cycle to convert from an in-house dictation system to vender-based dictation system
- Created an SQL database to provide forms and reports of patient records for the hospital-employed physician group, reducing the time in creating call schedules from two days to two hours
- Participated on the development team for the implementation of the Recovery Audit Contractors Program, which was a federal mandated initiative to identify potential cases of overbilling, which documented 100% compliance on launch

PUBLICATIONS

Qian L., Schmidt E. K., Scott R. L., and Homan K. D. (2016). An evaluation of Enterprise Resource Planning (ERP) systems in United Nations Organization using Critical Success Factor (CSF). L Proceedings of the 3rd ICOSTI Conference, Fukuoka Japan.

Scott R. L., Schmidt E. K., Zhao Y., and Homan K. (2015). Establishing and Managing Business-University Research Partnerships. Conference for Industry and Education Collaboration. Proceedings of the 2015 Conference for Industry and Education Collaboration. American Society for Engineering Education.

Homan K. D., Jones J. C. (2015). Legofycation of Education. Proceedings of the 2nd Six Sigma Conference, Edinburgh Scotland.

Homan K. D. (2014). Production Capacity Process Improvement in a High Mix Low Volume Company. Proceedings of ATMAE Conference St Louis Minnesota.