An Examination of Decision-Making Biases on Fourth Down in The National Football League

Weller Ross, B.S. Sport Management

Submitted in partial fulfillment of the requirements of the degree of

Master of Arts in Applied Health Sciences

(Sport Management)

Faculty of Applied Health Sciences

Brock University

St. Catharines, Ontario, Canada

Dedication

This thesis is dedicated to the loving memory of my uncle, David Ross, who passed away on April 14, 2016. Uncle David contributed to the field of sport management for 45 years, dating back to when he earned his undergraduate degree from the University of Tennessee and his master's degree from Ohio University.

He went on to have an extremely impressive career in arena management, which included serving as the president of the International Association of Venue Managers (IAVM) and being accorded the industry's most prestigious honor, the Charles A. McElravy Award, which signifies extraordinary contributions to the field. Most importantly, David was a loving uncle and a great mentor.

We love you, Uncle David. You will be missed.

Abstract

The recent developments in the field of sport analytics have given researchers the tools to examine an increasingly diverse set of topics within the world of sport in ways not previously possible (Alamar, 2013; Fry and Ohlmann, 2012). This study analyzes the decision-making processes of high level coaches under different contexts and then determines whether or not a specific subconscious psychological bias, known as the representativeness heuristic, caused the individual to make the choice they did.

Past empirical research has examined people's decisions in different contexts and, from those contexts, made inferences about how those individuals made their decisions and what errors in their decision-making processes could have led to their suboptimal choices (Kahneman and Tversky, 1979; Kobberling and Wakker, 2005; Tom et al, 2007; Tversky and Kahneman, 1992).

The representativeness heuristic explains that errors in people's judgment occur because their mind places too much emphasis on the current situation (new information) and not enough on the original odds (prior information). Previous researchers have been unable to separate the new and prior components of people's decision-making when studying real-world scenarios in a sport context (Carter and Machol, 1978; Carroll, Palmer, and Thorn, 1989; Carroll et al, 1989; Patel, 2012; Romer, 2006).

This research is different than the previous related research in that we utilize statistical models to gauge how people weight different information when making high-pressure decisions in sport. We hypothesize that coaches are disproportionately weighting new information against prior information when making decisions, and thus, yielding to the representativeness heuristic.

To test our hypothesis, we construct numerous Bayesian updating models to represent the impact of National Football League (NFL) coaches' decision-making on the likelihood of winning games. Utilizing a Bayesian approach enables us to keep the new and prior odds of winning the game separate, and thus, keep the two components of the representativeness heuristic separate. Regression analysis is then used with both of the components to directly test for the representativeness heuristic in NFL coaches' decision-making by estimating the effect each component has on the coaches' decisions. These estimates form the basis of our hypothesis tests.

Index terms— Bayesian, decision-making, representativeness, analytics, football

Acknowledgments

I would first like to thank my thesis advisor, Dr. Kevin Mongeon, for his guidance and solidarity throughout this entire research process. His knowledge, experience, and dedication to helping me succeed have far exceeded any of the hopes I had coming into graduate school. Over the course of these two years, I have learned more from him than I ever could have expected or anticipated. I am so thankful for having had this opportunity to learn under the tutelage of Dr. Mongeon. He has been the best mentor imaginable.

Additionally, I would like thank all of my committee members for the time and effort they spent helping me reach this point. I would also like to thank them, as well as my external examiner, for the input and insight they have contributed to this research.

Furthermore, I would like to convey the supreme gratitude I feel toward my parents, John and Cathy Ross, not only for their continuous love, support, and encouragement, but also for their endless confidence in my ability to succeed. I am also extremely grateful for the role my sister, Libby, and her husband, Matt, have played in my life. Having an older sister that I can look up to and always count on has been an invaluable asset to me.

I would also like to thank all of my friends for their support. Specifically, I would like to thank my best friend, Tavis Sartin, and my cousin, Bryan Ross, for inspiring me to pursue my dreams. Their undying loyalty and camaraderie has given me the courage to strive for success, never give up and continuously push myself to improve.

Finally, I would like to express the overwhelming appreciation I have for my wife, Gabrielle Ross. Throughout this process she has not only been my girlfriend, fiancée, and wife, she has also been my primary source for motivation and companionship. She has provided me with unwavering support and has believed in me more than even I have at times. Her presence in my life means more to me than I can possibly say.

Contents

List of Tables

List of Figures

1 Introduction

Do people consistently make suboptimal decisions? Past empirical research has examined people's decisions in different contexts and, from those contexts, made inferences about the errors in their decision-making and what could have led to their suboptimal choices. The two primary decision-making theories are Expected Utility Theory and Prospect Theory. Expected Utility Theory suggests that individuals make decisions by maximizing utility in terms of probabilities (Bernoulli, 1738; English trans. 1954; Von Neumann and Morgenstern, 1945, 2007). Prospect Theory argues that it is too difficult and computationally intensive for people to always maximize utility as Expected Utility Theory suggests, which can lead to bias (Kahneman and Tversky, 1979; Tversky and Kahneman, 1973, 1974).

Prospect Theory uses four well-established psychological biases (or heuristics) to explain why people make the decisions they do. These heuristics are loss aversion, representativeness, anchoring, and availability. The representativeness heuristic explains that errors in people's judgment occur because their mind over-generalizes previous similar events and places too much emphasis on the current situation (new information) and not enough on the original odds (prior information), rather than weighting both pieces of information appropriately. The heuristic causes relative insensitivity to prior information, the drawing of strong inferences from small sample sizes, a misconception of chance, the illusion of validity, and the underestimation in the likelihood of change in a trend. Previous researchers have been unable to separate the prior odds and conditional likelihood components of people's decision-making when studying real-world scenarios. This study uses a Bayesian approach to evaluate coaches' decisions, enabling us to keep the two components separate, which allows us to directly test for the representativeness heuristic.

Bayes' rule describes the normative decision-making process involving the optimal combination of two components: prior and new information. Our approach will observe these two cognitive components of decision-making processes on aggregated data. To examine their impact we will develop theoretical models of the Bayesian updating process by exploiting various aspects of in-game information related to the likelihood of event outcomes and real decisions made by coaches on fourth down. Finally, we will test whether or not the representativeness heuristic is reflected in the aggregated decision-making processes of coaches.

There have been studies to examine how individuals process information, but they use a laboratory setting to evaluate their subjects' decisions. They found that people underweight the prior probability of an event occurring (Grether, 1980, 1992), do not efficiently incorporate new information (Charness and Levin, 2005; Friedman, 1998), and that the influence of new information is the greatest when it contradicts the prior information (Ashton and Ashton, 1998). Charness et al. (2007) found that the subjects' error rates decreased when in groups versus in isolation, and concluded that systematic deviations from expected Bayesian updating "are due, in part, to artificial isolation imposed by experimental design" (p.147). In this project, we blend optimal coaching decisions and experimental approaches to examine how coaches' decision-making processes are, in aggregate, reflected in their fourth down decisions.

In the National Football League (NFL) a head coach is required to make difficult decisions in a variety of situations throughout the course of a game. One of these situations is when his team is facing a fourth down on offense. This study evaluates the decisions of professional football coaches in those fourth down situations. These are important situations because a team on offense gets four attempts (downs) to gain at least a net total of 10 yards while the opposing team's defense tries to stop them. If they are successful then they are awarded a new set of four downs to gain another 10 yards, and this pattern continues until they score, time runs out, or they fail to gain the required 10 yards, in which case they relinquish possession of the ball to the opposing team, which then attempts to move the ball in the opposite direction under the same process.

On fourth down a team can attempt to gain a first down, or kick the ball by either attempting a field goal for three points or punting to the other team, both of which would result in the opposing team gaining possession of the ball. It is much more common for a team to score when it is on offense, and thus, a fourth down is an important situation for coaches to make an optimal decision because it is the last opportunity his team has to maintain possession of the ball by gaining a new set of downs.

Choosing the sport of American football, and specifically within the context of the NFL, to evaluate an individual's decision-making processes for the representativeness heuristic was done because it provides a circumstance in which the decision-maker (head coach) and their decisions are frequently and regularly making an impact on the likelihood of the potential game outcomes. Additionally, evaluating fourth down decisions specifically, introduces a circumstance in which we know that the head coach is the one making the final decision on what to do, which also lends itself to the research of Charness et al. (2007) when they found that individuals' error rates decrease when in groups rather than isolation. While coaches are not literally isolated when they make decisions on fourth down, they are figuratively in isolation as the lone person responsible for making the decision.

Similarly, the findings of Ashton and Ashton (1998) also make fourth downs in the NFL an intriguing circumstance for examining the impact of the representativeness heuristic on an individual's decision-making process. Their results showed that the influence of new information is the greatest when it contradicts the prior information, and that this can lead to suboptimal decision-making (Ashton and Ashton, 1998). Specifically, this can cause the decision-maker to be more susceptible to biases, such as the hot-hand fallacy or gambler's fallacy, which are both byproducts of the representativeness heuristic (Sundali and Croson, 2006). As mentioned previously, this heuristic causes relative insensitivity to prior information, the drawing of strong inferences from small sample sizes, a misconception of chance, the illusion of validity, and the underestimation in the likelihood of change in a trend—all of which can come into play when a coach is making a decision on fourth down. Thus, the findings in this research provide support that previous criticisms of the claims that have been made based on experimental research are potentially overstated.

This study determines whether or not coaches are yielding to the representativeness heuristic when they are making these fourth down decisions. This examination is conducted by first uncovering which of the options on fourth down is the optimal decision in different circumstances (game-states) based on its impact on the odds of winning the game. Second, we measure the relative weight used by coaches for the new and original information when they are making their decisions.

2 Literature Review

There is a large body of previous research which has examined decision-making in the NFL. A number of those researchers have examined what the optimal decision is on fourth down. All of the researchers that examined fourth down decision-making (Carroll et al., 1989; Carter and Machol, 1971, 1978; Patel, 2012; Romer, 2002, 2006) have found that teams/coaches act too conservatively by opting to kick more often than they should. Where this study uses a win probability model, previous researchers used expected points models to evaluate the impact of fourth down decisions on a game. An expected points model looks at different situations in the past to see which team was the next team to score and how many points they scored. The difference between each team's average is used as the amount of points a team can expect to score (or allow) in a similar situation.

There are several notable win probability models that are well-respected. For example, Brian Burke made an in-game win probability model using regression analysis with Markov model interpolations. Frank Frigo and Charles Bower also made an in-game win probability model using a program they created in 2005 called 'ZEUS Football' which utilized the techniques of Monte Carlo simulation. The website FiveThirtyEight made a pre-game win probability model by implementing the approach used for the Elo rating system, which was originally developed to provide rankings for chess players (Elo, 1978). Similarly, the website Pro-Football-Reference introduced their own in-game win probability model in 2013, building upon work previously done by Stern (1991) and Winston (2012). However, none of these win probability models were developed within the realm of academia, which detracts from their validity, despite how well-respected they are outside of the peer-reviewed field.

Carter and Machol (1971) created the first expected points model for the NFL, using data from the games of the first half of the 1969 season. They based their research on finding the expected points for a team with a first down at different field positions. Carter and Machol (1978) used data from the 1971 season to build off of their previous research (1971) by finding out how the expected points value would be impacted by the decision to punt, attempt a field goal, or go for it on fourth down while on the opponent's side of the field. They found that, in those situations, coaches were choosing to kick field goals with a much higher frequency than they were deciding to go for it, even in situations where the expected points value would be greater by going for it. They believed that coaches were failing to take into account "the negative value imposed on the opposition team when they are given possession of the ball in the shadow of their own goal post."

Carroll, Palmer, and Thorn (1988) used data from the 1986 NFL season to construct an expected points model with both a similar method and similar findings to those of Carter and Machol. They found that when a team is facing a fourth down with six yards to gain, or less, it was hardly ever the best choice to attempt a field goal despite the fact that teams were kicking the ball the majority of the time. There are some key limitation to the research done by Carter and Machol (1971 and 1978) and Carroll, Palmer, and Thorn (1988) in that they assumed linearity across the different field positions and they did not evaluate their results for statistical significance. They also admitted that they believed the data sets they were working with were too small. Most importantly, they did not account for time remaining or score margin.

In a well-cited paper, Romer (2006) examined first quarter NFL play-by-play data from 1998-2000, resulting in 11,112 observations. He used an instrumental variable approach to estimate the expected points value of having a first down with 10 yards to go at each yard line. Romer used these results to estimate how many yards to go would have an equal value for kicking and attempting a first down across each yard line. He then decided what the optimal fourth down decision would be based on if the number of yards to go was more or less than the corresponding critical point for the yard line they were on. Romer then compared his optimal fourth down decisions to coaches' actual decisions and found that of the 1,068 fourth downs in which his analysis indicated that teams should attempt a first down, they instead kicked 959 times (89.794 percent), prompting him to state, "examination of actual decisions shows systematic, clear-cut, and overwhelmingly statistically significant departures from the decisions that would maximize teams' chances of winning."

While Romer's method allowed for a non-linear expected points model and his results showed statistical significance, there were still some limitations to his study. The research included a larger data set, but he largely negated this improvement by only using data from the first quarter of each game. He does this to get around having to account for the score margin and time remaining, "to avoid the complications introduced when one team is well ahead or when the end of a half is approaching, I focus on the first quarter."

Patel (2012) made an expected points model using data from the 2007-2011 seasons to evaluate coaches' fourth down decisions. To avoid accounting for score margin and time remaining, Patel cut down his data as well, developing his expected points model using only plays that occurred in the first or third quarter and with a score margin of 10 points or less. Patel also used Brian Burke's win probability calculator (http://wp.advancednflstats.com/winprobcalc1.php) to examine the coaches' decisions on every fourth down play from 2007-2011 that took place in the second or fourth quarter and were within 30 yards of the opponent's end zone. Patel found that "the behavior of coaches in the National Football League on fourth downs does not align with the behavior that optimizes a teams probability of success. This fact remains true even though fourth-down decisions are relatively simple." Patel went on to mention that prospect theory and loss aversion could be part of what is causing coaches to make these suboptimal decisions, "coaches might value losses of a play higher than they would the corresponding gains of a play and so they might be calling conservative plays."

Common findings among these papers suggest that other unknown factors, beyond those that directly impact the situation, were influencing the coaches' decision-making. However, none were able to definitively discern the underlying source of the bias or biases. Romer (2006) states "there is little evidence about whether conservative behaviors arise because individuals have nonstandard objective functions or because they are imperfect maximizers." Romer went on to mention it could be that those involved in the decision-making process are being profit-maximizing rather than win-maximizing, or that the decision makers are systematically imperfect maximizers. It could also be that they perceive the cost of losing as the result of a failed gamble as greater than the cost of losing from playing it safe, which could be influenced by any number of factors including job security, fan support, the media's reaction, players' trust, or internal/personal feelings, such as experiencing more regret (Romer, 2006).

The researchers had to make inferences from the contexts they analyzed to determine what could be causing the coaches to make suboptimal decisions. However, due to their methodologies, none of them were able to provide conclusive evidence to verify or validate their conclusions. This is because, despite their varying methods being appropriate to determine whether or not coaches are making suboptimal decisions, they were not conducive to examining the information processing procedure of the coaches. This means that while their methodologies did allow them to determine *if* coaches were making suboptimal decision, it prevented from finding out *why* they were making suboptimal decisions.

Furthermore, none of the previous researchers accounted for team quality in their analyses, but instead worked under the assumption that the two competing teams were evenly matched or average. This study employs Bayes' rule, which provides us with the flexibility to account for team quality and keep the prior odds separate from the conditional likelihood to assess whether or not the representativeness heuristic is a factor that is causing coaches to make suboptimal decisions.

3 Methodology

3.1 Theoretical Approach

Bayesian updating models and a latent variable approach are used to test for the representativeness bias. Specifically, a Bayes' rule formulation is used to represent the in-game likelihood of winning based on pre-game probabilities and in-game information (game-state). Then regression analysis estimates the weights of the information using the prior probability and conditional likelihood components to test for representativeness in NFL coaches' decision-making. Bayes' rule illustrates the operation of combining those two components to arrive at the optimal conclusion as Expected Utility Theory would recommend. We will use Bayes' rule to examine the decisions that NFL coaches are making on fourth down and to determine whether or not they are guilty of the representativeness heuristic.

We assume NFL coaches have the following options for their 4th down decisions: 1) first-down attempt (FDA), 2) field goal attempt (FGA), or 3) punt (PNT). If a team achieves a first down they retain possession of the ball and are awarded a new set of downs. If they attempt a first down and fail, they relinquish possession of the ball to the opposing team on the yard-line at which they were stopped. Teams that successfully kick a field goal are awarded three points and then kickoff to the other team. Missed field goal attempts result in the opposing team taking possession of the ball with their starting field position being the point from which the ball was kicked. A punt gives the opposing team possession of the ball on the yard line at which the punt goes out of bounds, is downed, or where the return man is stopped. If the punt goes into the end zone, the opposing team takes possession of the ball at their own 20-yard line. It is unlikely a team retains possession of the ball when they choose to punt or attempt a field goal. They only retain possession if the opposing team commits a penalty that awards them the necessary yardage for a first down, or if the opponent fumbles (or touches) the ball, which is then recovered by the kicking team.

The likelihood, or odds (o), that the g^{th} game is eventually won by team c given each potential decision, d, and the current game-state ($s_{g,t}$) at time t can be expressed as

$$o^*(c_{g,t}|d_{g,t}, s_{g,t}) = \hat{o}(c_{g,t}|s_{g,t}) \times \hat{o}(d_{g,t}|c_g, s_{g,t}) \tag{1}$$

Equation (1) is Bayes' rule represented in odds form. The left-hand side term of Equation (1) denotes the posterior odds of a game win given the decision and game-state. Note, it is straightforward to represent odds as the ratio of probabilities: $\frac{p^*(c_{g,t}|d_{g,t},s_{g,t})}{p^*(\bar{c}_{g,t}|d_{g,t},s_{g,t})}$. The first right-hand side term is the in-game odds of a win prior to the decision, and the second term is the likelihood ratio, or inverse conditional decision odds.

Equation (1) can be transformed into log-odds and expressed as:

$$o^*(c_g|d_{g,t},s_{g,t}) = \beta_0 + \beta_1 ln(\hat{o}(c_{g,t}|s_{g,t})) + \beta_2 ln(\hat{o}(d_{g,t}|c_g,s_{g,t})) + \varepsilon_{g,t} \tag{2}$$

$$\text{where } o(c_g|d_{g,t}) = 1 \text{ if } o^*(c_{g,t}|d_{g,t}) \geq 1.0 \text{ and 0 otherwise} \tag{3}$$

Finally, the in-game win odds, $\hat{o}(c_g)$ is also made to be estimated throughout the progression of games with Bayes' rule.

$$\hat{o}(c_{g,t}|s_{g,t}) = o(c_{g,t=0}) \times o(s_{g,t}|c_g) \tag{4}$$

Equation (4) represents the in-game odds (o) team c has of winning game g given the game-state (s) at time t. The first right-hand side component, $o(c_{g,t=0})$, is the pre-game (prior) odds (o) of team c winning the game ($t = 0$). The second right-hand side component, $o(s_{g,t}|c_g)$, is the conditional likelihood (o) of the game-state (s) occurring at time t given that team c wins the game.

Equations (2) and (3) forms the basis of the representativeness hypothesis testing procedure, showing that suboptimal decision-making exists. Equation (4) provides the necessary in-game odds (new information), while Equation (5) provides the separate conditional likelihood information for the decision based on historical data (original information), which are the two necessary components to conduct a test for representativeness.

3.2 Empirical Approach

Prior to estimating the unknown parameters of Equation (1) it is necessary, for empirical purposes, to estimate the in-game win odds prior to the decision, $\hat{o}(c_{g,t}|s_{g,t})$, and the likelihood ratio, $\hat{o}(d_{g,t}|c_g,s_{g,t})$. To do so, we use a multinomial logistic regression to estimate the odds of each decision being chosen by the coach. This results in three different conditional likelihoods and thus, a different posterior for each of the three choices, which allows us to evaluate the decision itself rather than the outcome from the decision. This is important because it is possible for a coach to make a good decision that has a bad result and vice versa. The posterior odds can then be compared to the in-game pre-decision win odds for each option and the decision that has the most favorable impact on the team's win odds is the optimal decision.

Note, all NFL games designate one team as the home team and the other team as the away team, even for games that are not played at one of their home stadiums. We use c as the team designation for the decision-makers (the offense) and $\bar{c}$ as their opponent (the defense). Similarly, we use h as the home team and a as the away team. This is important because we represent all game outcomes in the data relative to the home team (h) for Equations (4), (6), and (7) while we represent all game outcomes in the data relative to the team on offense (c) for Equations (1), (2), (3), (5), and (8). (ie. If $c_g = h_g,$ then $\bar{c}_g = \bar{h}_g = a_g$ where h_g and a_g are the home and away team respectively.)

3.2.1 Likelihood Ratio Estimation Procedure

We use the following regression equation to estimate the odds of each decision (d) being made, given the game-state components (θ) and game outcome (c_g).

$$\hat{o}(d_{g,t}|c_g, s_{g,t}) = F^{ML}(X^{'}_{g,t}\theta + \varepsilon_{g,t})|c_g \tag{5}$$

On the left-hand side, the d term is a categorical variable representing each of the three potential decisions. Those three options are represented as d^{FDA} = a first down attempt, d^{FGA} = a field goal attempt, and d^{PNT} = a punt. On the right-hand side $X^{'}$ represents the matrix of variables that the game-state is comprised of: the score margin, time remaining, timeouts remaining, field position, and yards to go for a first down, as well as a time-adjusted value of the bookmakers' over/under (total points scored by the two teams combined) for the game. It should be noted that while the current down is one of the variables that is part of the game-state, it is not included in this matrix because the subset of data used for Equation (5) contains only fourth down plays.

3.2.2 In-game Win Odds Estimation Procedure

The in-game odds of a win prior to the decision is estimated through the win odds model represented by Equation (4). To estimate those odds, we first need to estimate the pre-game odds of winning and the odds of the game-state given the outcome of the game, to be used as the prior and likelihood ratio components, respectively.

The closing bookmaker point spreads are transformed into game outcome win odds with a logistic regression and used as a proxy variable for the pre-game odds of winning, which are represented as

$$o(h_{g,t=0}) = F^L(\psi_0 + \psi_1 X^1_{g,t=0} + \varepsilon_{g,t}) \tag{6}$$

where $o(h_{g,t=0})$ is the odds (o) of the home team (h) winning the game (g) before it begins ($t = 0$), while $X^1_{g,t=0}$ represents the game's closing point spread with respect to the home team. Figure 1 illustrates the reliability of using point spreads as a proxy variable for the pre-game odds of winning.

The likelihood ratio of the game-state (s) at time t is estimated with a multinomial logistic regression and is represented as

$$o(s_{g,t}|\hat{h}_g) = F^{ML}(X^{'}_{g,t}\lambda + \varepsilon_{g,t})|h_g \tag{7}$$

where $X^{'}$ again represents a matrix of variables: time remaining, timeouts remaining, field position, down, yards to go, possession, and the time-adjusted closing bookmaker over/under data for the total points scored.

This model, coupled with Equation (6), captures the relevant information that a coach should be taking into consideration when making a decision on fourth down. All of the variables that the game-state is comprised of are accounted for, and by also including the closing bookmaker data we are able to account for not only team quality, but other additional factors such as weather and stadium effects as well.

3.2.3 Hypothesis Testing Estimation Procedure

A logistic regression equation is used as the test for representativeness, which is expressed as

$$o^*(c_g|s_{g,t}, d_{g,t}) = F^L(\varphi_0 + \varphi_1 \ln \hat{o}(c_{g,t}|s_{g,t}) + \varphi_2 \ln \hat{o}(d_{g,t}|c_g, s_{g,t}) + \varepsilon_{g,t}) \tag{8}$$

where the null hypothesis (H_0) is that the parameter estimates (φ_1 and φ_2) for the prior odds component and likelihood ratio component are equal to each other, suggesting that the coaches are equally weighting original information and new information in their decision-making process.

For the purposes of this study, new information is being defined as the team's current in-game win odds, while original information is being defined as the odds of a coach making the decision that he did given that his team ended up winning the game, which is generated from historical data.

4 Data

This analysis uses play-by-play data from each regular season game over a 12 year span (2003 - 2014). The six odd-numbered years are used to estimate the in-game win odds, while the six even-numbered years are used to test the decision-making hypotheses. The data set goes back to 2003 because that allows for a sufficiently large dataset without going back to 2002, which was the most recent expansion year for the league. Going through the 2014 season without including the 2015 season was to account for the fact that there was a rule change for the 2015 season that moved back extra point attempts, which impacts how likely a team is to successfully convert a point-after-try (PAT) following a scored touchdown.

For each play/observation there is information containing the decision and associated game-state in which the play took place. This includes the closing bookmaker information, identifiers for the home team and away team, the score of the game, time remaining, field position, down, yards to go, play type, timeouts remaining, and the outcome of the play, as well as an indicator variable for whether or not the team won the game.

To estimate the in-game win odds, the data is organized into panel data by the season, game and play. This results in 468,699 game-play observations with all game outcomes relative to the home team. For the decision-making model the data is filtered to only include fourth down plays with all game outcomes relative to the possession team, resulting in 45,559 (9.7 percent) game-play observations. First, second, and third down were 36.1, 27.2, and 17.3 percent respectively. The remaining observations were kickoffs (6.4 percent) and extra points (3.2 percent).

The summary statistics for the conditional likelihood component of the optimal decision model are provided in Table 1 across each decision option (first down attempt, field goal attempt, and punt) for the variables that make up the game-state. This includes yards to go, absolute yard line, minutes remaining, score margin, in-game win probability and pre-game win probability with respect to the offensive team, as well as the bookmaker data on point totals, the number of timeouts remaining for each team, and an indicator variable for the home team.

In terms of the fourth down plays, 64.02 percent (29,186) were punts, 23.21 percent (10,583) were field goals, and 12.70 percent (5,790) were first down attempts. The average yards to go was 7.66 with a standard deviation of 5.66. Of the 5,790 first down attempts, 3,940 (68.05 percent) occurred when the offensive team was trailing (21.61 percent when leading) and 56.29 percent were in the fourth quarter (third quarter: 14.91 percent, second quarter: 17.13 percent, first quarter: 11.68 percent). This indicates that teams are typically attempting to gain a first down when they are already in a losing situation. This is occurring despite the fact that 46.7 percent of the first down attempts were successful. That is a much higher success rate than the combined average for the first three downs (27.47 percent) or any of the other three downs individually, the highest of which was on third down at 38.90 percent (first: 19.53 percent, second: 30.75 percent). Similarly, teams that were "underdogs" (meaning that bookmakers predicted them to lose the game) attempted first downs 13.8 percent of the time while teams that were favored to win were at 11.5 percent.

Losing teams attempted a first down 16.22 percent of the time while winning teams attempted a first down 8.80 percent. Teams with an in-game win probability of less than 0.50 attempt a first down 17.21 percent of the time while teams above 0.50 are at 7.88 percent. This is further indication that teams are deciding to attempt a first down with a higher frequency when they are already likely to lose. The success rate for the losing teams was 41.80 percent while the winning teams' was 59.19 percent. This indicates that despite the fact that winning teams successfully convert their first down attempts with nearly 150 percent the frequency of the losing teams, they attempt first downs approximately half as often. This indicates that rather than choosing to attempt a first down based on when it is the optimal decision, coaches are instead choosing this option out of desperation.

The components of the win probability model can be seen in Tables 2, 3 and 4, which collectively present the summary statistics for the prior and conditional likelihood. Table 2 presents the mean, standard deviation, maximum, and minimum values for the prior probability (with respect to the home team) for games in which the home team won, lost and tied. The summary statistics for the variables in the conditional likelihood are presented in Tables 3 and 4 for the offense and defense respectively. The data is shown across wins, losses and ties, and includes time remaining, down, yards to go, absolute yard line, and score margin, all with respect to the home team. The number of timeouts remaining for each team and the predicted point total are provided as well. Tables 3 and 4 are separated by offense and defense to account for a possession indicator variable that is also included in the model.

5 Results

5.1 Theoretical Models

5.1.1 Decision Model

The results from Equation (1) are reflected in Figure 2, which illustrates the change in an offensive team's win probability based on the decision they make on fourth down with respect to how many yards they need to gain a first down. Figure 2 was generated using the fourth down plays with 10 or fewer yards to go that occurred when the score margin was less than or equal to eight, and took place on or between the defensive team's 33- and 40-yard lines.

These criteria were chosen because they represent a situation in which all three fourth down options are viable choices while the score is still close. When the two teams' scores are within eight points of each other, it is also referred to as the teams being within one possession of each other, due to the fact that eight is the maximum number of points a team is able to score on a single possession. Being within this section of the field would require a long field goal attempt, while punting would be unlikely to net the team very many yards, and attempting to gain a new set of downs and failing would leave the opposing team with favorable starting field position.

This type of situation is known as being in no man's land, which is a vague term that refers to a section of the field, where the exact yard lines are unspecified. For the purposes of this study it was decided that identifying a range where the three different options are similar both, in terms of how many times they have been chosen, and in terms of their expected likelihood of success, would be best for analyzing the optimal decisions across yards to go. To be considered successful field goal attempts had to be made, first down attempts had to result in a new set of downs for the offense, and punts had to result in the opponent starting their possession inside of their own 11-yard line.

From the 36- and 37-yard lines (the middle of the area specified above), punts were successful 190 out of 355 times (53.52 percent), field goals were successful 111 out of 202 times (54.95 percent), and first down attempts were successful 145 out of 272 times (53.31 percent). Three yards were added on to each side of the range to give us the area from the 33-yard line to the 40-yard line. Within this range all field goal attempts would be of 50 or more yards, and all punts that resulted in a touchback would net the team just 20 or fewer yards, and the expanded range resulted in only slight changes of the success rates. Punts were successful 53.04 percent of the time and field goal attempts were successful 58.30 percent of the time, while first down attempts with 10 or fewer yards to go were successful 51.16 percent of the time.

As can be seen in Figure 2, with about seven yards to go is where the critical point is for deciding between the choice to attempt to gain a first down and the choice to kick a field goal. This is consistent with the findings of Romer (2006) when his results showed that "after mideld, the gain from kicking falls, and so the critical value rises. It is 6.5 yards at the opponents 45 and peaks at 9.8 on the opponents 33." It is also noteworthy that on average punting in this situation is never the optimal decision and is only the second-best option once there are 10 yards to go. Despite this, punting is the most frequently chosen decision. Of the 3,535 times teams were facing a fourth down on this section of the field, coaches chose to punt 1,548 times (43.79 percent), which is 706 more times than they chose to kick a field goal (842 or 23.81 percent) and 403 more times than they chose to attempt a first down (1,145 or 32.39 percent).

Additionally, choosing to punt under these conditions decreases a team's win probability across all yards to go. Deciding to punt with one yard to go would result in a change to the possession team's win probability of approximately -0.06, while punting with 10 yards to go results in a -0.02. Attempting a field goal starts at about -0.01 with one yard to go and then hovers around a 0.00 change in win probability from two to seven yards to go, at which point it starts to rise to its maximum increase in win probability of 0.02. Lastly, attempting a first down starts at its maximum with an increase in win probability of approximately 0.08 when there is one yard to go, then decreases until its critical point with field goal attempts and the x-axis (a 0.00 change to the win probability) at about seven yards to go before reaching it's minimum of approximately -0.03 when there are 10 yards to go.

With seven yards to go being the point at which it becomes no longer the optimal decision to attempt a first down, and instead is where it becomes equally beneficial to kick a field goal, it would be expected for coaches to choose a first down attempt with a higher frequency when they are less than seven yards away from gaining a first down. However, it is only with one or two yards to go that a first down attempt was the most common choice; at three yards to go coaches selected to punt 43.75 percent of the time, while only choosing to kick a field goal or attempt a first down 22.73 and 33.52 percent of the time respectively. For each of the other number of yards to go, coaches chose to punt more than 51.41 percent of the time, while neither of the other two options were ever selected with a frequency of more than 33.80 percent.

These trends hold true when evaluating the optimal decision across absolute yard line. Where Figure 2 illustrates the optimal decision across yards to go, Figures 3 and 4 illustrate when it is the optimal decision to attempt to gain a first down rather than to punt or kick a field goal across absolute yard line. Figure 3 uses plays within the 40-yard line to compare first down attempts and field goals, while Figure 4 examines plays from the defensive team's 40-yard line and beyond to compare first down attempts and punts.

Figure 3 was created using the data from fourth down plays where the offense had 10 or fewer yards to go for a first down, in the same way that Figure 2 was generated. Unlike, Figure 2 however, Figure 3 includes data from all plays in which the score margin was within two possessions (between -16 and 16) rather than one possession. This was done to ensure that there would be plenty of situations where making a field goal would be of high importance for the offense. For example, if a team is trailing by 10 points then they would need both a touchdown and a field goal to tie the game. Similarly, if a team is leading by 14 points and they make a field goal, they would not only be increasing their lead by three points, but they would then also be leading by three possessions, as there would be no way for the opposing team to score 17 points in just two possessions.

The results show that, on average, it is better to attempt to gain a first down rather than attempt a field goal inside of the opponent's 40-yard line, with the change in win probability being higher for first down attempts across all of the yard lines. The two estimates are closest around the 10-yard line because as a team gets closer and closer to the goal line (the 0-yard line) it becomes easier and easier for the kickers to make field goals while it becomes harder and harder for a team to gain yards (and score a touchdown) due to their play-calling options becoming more limited and the opposing team having less area to worry about defending. This is reflected in Figure 3 with the two lines drawing close to each other at about the 30-yard line and staying close until just after they nearly meet at about the 10-yard line.

Outside of this 20-yard area from the 30-yard line to the 10-yard line is where we see the biggest differences between a first down attempt and a field goal attempt. Going from the 30-yard line to the 40-yard line it is logical for the field goal attempts to decrease as that is when it starts to approach the limits of most field goal kickers' range, and therefore it becomes more beneficial to attempt a first down, which is reflected in Figure 2 as well. Inside of the 10-yard line is when it starts becoming more likely that the offensive team would be able to score a touchdown on (or subsequently to) a first down attempt because the defense would likely be unable to prevent the offense from scoring for four more plays. The result of this can be seen in Figure 3 around the five-yard line, where the increase in the separation between the two lines starts to become more obvious. These findings are again similar to those of Romer (2006), who wrote that his "analysis implies that once a team reaches its opponents 5, it is always better off on average going for it."

Teams with a score margin from -16 to 16 faced a fourth down on or inside of the opponent's five-yard line 1,723 times. In this situation they opted to kick a field goal 1,253 times (72.72 percent) while only attempting to gain a first down 470 times (27.28 percent). Teams on or inside of the 10-yard line (n = 3,279) were even more lopsided in their decisions, choosing to kick field goals 79.93 percent (2,621) of the time. Between the 10- and the 30-yard lines teams attempted field goals 4,123 times out of 4,870 occurrences (84.66 percent) while attempting to gain a first down 746 times (15.32 percent) with the one remaining play being a punt. Of the 3,129 fourth downs teams faced from the 30-yard line to the 40-yard line, as they reached the fringe of field goal range, they actually chose to attempt a first down (1,026) nearly as many times as they chose to attempt a field goal (1,129). However they also chose to punt (974) nearly that many times as well, which, as was discussed above with respect to Figure 2, is not the optimal decision in these situations.

Figure 4 was created by further separating the field into two more sections. For the half of the graph that spans from the 40-yard line to the 70-yard line, it still uses the data from fourth down plays with 10 or fewer yards to go, but in situations where the score margin was between -11 and 11, while the portion of the graph between the offensive team's 30-yard line and their own goal line (from the 70 to the 100 in terms of absolute yard lines) uses data from games within two possession, as before with Figure 3. Using score margins between -11 and 11 allowed for a higher ratio of situations in which a team in the middle of the field would be simply trying to get into field goal range. This is due to the fact that teams trailing by 9, 10, or 11 points could tie the game or take the lead with a touchdown and a field goal, while teams trailing by more than that would need two touchdowns.

The dashed line representing the change in win probability as a result of deciding to punt stays close to zero across all of the yard lines. This is reflecting an underlying effect from teams choosing to punt from their own half of the field (absolute yard lines larger than 50) the large majority of the time, regardless of how many yards they are from a first down, and whether or not they are leading or trailing in the game. They tend to only choose a first down attempt when they are in a desperate situation late in the game, which is also reflected in Table 1 and was discussed previously in the Data section. This impacts the output from the conditional likelihood components of Equations (1) and (4), which means the in-game win probabilities on this side of the field are already accounting for the fact that there is a high likelihood that the team is going to punt. This improves the accuracy of the estimated win probabilities and thus, when the team actually decides to punt, the resulting change in their win probability is minimal.

Additionally, starting around the 45-yard line there is a decrease in the change in win probability for choosing to punt. This is because the likelihood of making a field goal is increasing and the cost of a failed first down attempt is decreasing, while the maximum potential benefit of a punt is decreasing as well. It could also be due, in part, to the team being closer to no man's land, where it is no longer overwhelmingly likely that the team will opt to punt, and therefore, there is more of an impact to the change in win probability as the result of a team deciding to punt.

The solid line on Figure 4 reflects the change in win probability for if a team were to choose to attempt a first down rather than punt. For absolute yard lines higher than 80 (the possession's team 20-yard line) there is a negative impact on the win probability. This is logical, as a failed first down attempt in this section of the field would give the opposing team possession of the ball in a situation where they would already be very likely to score. Romer (2002) found that, "even on its 10-yard line - 90 yards from a score - a team within 3 yards of a first down is better off on average going for it" and he went on to say "although these findings contradict the conventional wisdom, they are quite intuitive." It makes sense for Romer to mention the 10-yard line in this context (rather than the 20-yard line), because he is referring to teams with three yards to go (or less) for a first down, while the majority of the plays used for generating Figure 4 had more than three yards to go to gain a new set of downs, with an average of 8.43 and median of seven.

For the next few yard lines on the graph (between the 77- and 80-yard lines) choosing to attempt a first down continues to be approximately equally as beneficial as deciding to punt. However, at about the team's own 24-yard line (absolute yard line of 76) it starts to have a larger positive impact on the team's win probability when deciding to attempt a first down instead of punting. In the graph Romer (2002) is referring to in the above quote, it appears that around the 25-yard line, a team with five or fewer yards to go for a first down should, on average, attempt to gain a first down rather than punt. My findings once again align with his as it is around the 25-yard line where Figure 4 indicates that it becomes more beneficial to attempt a first down, and in the data used to generate Figure 4 the median yards to go for fourth downs at that point on the field is five yards to go.

As alluded to before, the ratio of times teams have chosen to punt compared to the number of times they have chosen to attempt a first down is extremely disproportionate on the section of the field covered in Figure 4. Fourth downs have occurred 21,345 times at absolute yard lines greater than 50 while the score margin was within two possessions. In these situations teams have decided to punt on 20,276 (94.99 percent) of those occasions while the other 1,069 (5.01 percent) were first down attempts. From the 40-yard line to the 50-yard line the decisions are less lopsided but still heavily unbalanced with 4,176 (85.28 percent) punts, 692 (14.13 percent) first down attempts, and 29 (0.59 percent) field goal attempts.

5.1.2 Win Odds Model

The results of the win odds model shown in Equation (4) are illustrated in Figure 5, which provides the results across time remaining in the game. Specifically, Figure 5 shows the results in the form of probabilities (rather than odds) with respect to the team on offense across five different score-states. Figure 5 was generated using the data from teams that had possession of the ball with a first down on their own 20-yard line (80 yards from scoring), and 10 yards to go for a new set of downs.

These specifications were chosen because out of all the possible combinations of down, distance, and field position, this particular combination represents the most common situation in football with 11,423 occurrences in this data set, which is nearly four times more common than the situation (first and 10 at the 69-yard line) with the second-most occurrences (3,079). This is due to the fact that the rules in the NFL during this time period caused touchbacks (when a kickoff or punt lands in or beyond the end zone) to result in a team taking possession of the ball at their own 20-yard line (for the 2016 season, the NFL is doing a one-year trial where this will be moved to the 25-yard line) and every possession starts with a first down and 10 yards to go, barring those that begin within 10 yards of the opposing team's end zone. This is also why the second-most common situation takes place exactly 11 yards further down the field.

This situation is used to examine the in-game win probabilities of the following five score-states:

1. leading by two or more possessions (score margin of 9 or more)
2. leading by one possession (score margin from 1 to 8)
3. tied (score margin equal to 0)
4. trailing by one possession (score margin from -8 to -1)
5. and trailing by two or more possessions (score margin of -9 or less)

The resulting graph (Figure 5) is extremely logical. One would expect that, holding all else equal, leading by two possessions is always better than leading by one possession, which is always better than being tied, and so on. Figure 5 reflects that thought process with the highest win probability always being for teams leading by two possessions and the lowest always being for teams trailing by two possessions, with the other three score-states fitting in as expected without any of the five lines ever intersecting with one of the others.

Figure 5 is also logical with respect to time remaining. When there is less time remaining in the game, there is less time for the trailing team to come back, and less time that the leading team would need to protect that lead. Thus, it would be expected that as time goes on in the game the likelihood of winning would increase for teams with the lead, and decrease for teams that are trailing. Similarly, if the game is tied, it is reasonable to expect the team on offense to be slightly more likely to win the game based on the fact that the team with possession of the ball has a better chance to score than the team on defense. Again, these rationales are reflected in Figure 5 as the in-game win probabilities increase over time for teams with the lead, while they decrease for teams that are trailing. And with the data being in respect to the team on offense, the line for teams that are tied starts at slightly above a 0.50 in-game win probability, and very gradually increases to just under 0.55 before taking a slight decline at the very end, which reflects the fact that the team would not have enough time to score and would therefore result in the game going to overtime, which causes the graph to regress back toward the 0.50 win probability.

The reliability of the in-game win odds model, Equation (4), was tested by using its posterior to predict actual game outcomes with a regression model. This regression was run using a validation set comprised of even-numbered years (2004, 2006, 2008, etc.) while the in-game win odds model itself was developed using the odd-numbered years.

$$c_g = F^L(\xi_0 + \xi_1 \ln \hat{o}(c_{g,t}|s_{g,t}) + \varepsilon_{g,t}) \tag{9}$$

Equation (9) returned a parameter estimate of 0.9925 for ξ_1 with a standard error of 0.05 and a p-value of less than 0.001, indicating statistical significance. A positive parameter estimate for ξ_1 demonstrates that there is a positive correlation between the estimated in-game win odds and the actual outcome of games. A parameter estimate of 0.9925 allows us to say that with a 10 percent increase in the estimated in-game win odds, the probability of a win increases by 0.099 (or 9.9 percent).

5.2 Empirical Models

5.2.1 Pre-Game Win Probability Model

The first component of the in-game win odds model is the pre-game win probabilities, which serve as the prior component of Equation (4). The pre-game win probabilities are estimated with Equation (6), the results of which are illustrated in Figure 1. An important note here is that a negative value for a point spread means that team is projected to win the game while a positive value means they are projected to lose, and that adding a team's point spread to their predicted final score will equal their opponent's predicted final score. For example, if a team is projected to win with a final score of 35 to 20, then their point spread would be -15, because their projected score (35) needs to add to their point spread to equal their opponent's projected score (20). Within that same example, their opponent's point spread would be +15, and would be calculated by following the same logic.

To convert these point spreads into win probabilities, we utilize a logistic regression (Equation (6)) to predict wins from point spreads with respect to the home team. The resulting parameter estimate was -0.1463, with a standard error of .007 and a p-value less than 0.001, which indicates statistical significance. It makes sense for the point spread's parameter estimate to be a negative value because, as explained above, the further negative a point spread is, the more the team is projected to win by. To demonstrate what Figure 1 is telling us, consider the example from before; if we assume that the team with a point spread of -15 is the home team we see on Figure 1 that this equates to approximately a 0.9 pre-game win probability, while if a home team had a point spread of -3 or +6, their pre-game win probability would be about 0.6 and 0.3, respectively.

5.2.2 Conditional Likelihood of Decision Model

The results of the decision model's conditional likelihood component (Equation (5)) are shown in Table 5 with respect to the multinomial logistic regression's omitted option, which is first down attempts. The results are shown for both wins and losses. All but three of the estimates are statistically significant at the .05 level, with all but six showing significance at the 0.005 level.

The parameter estimates for absolute yard line and yards to go are consistent with information that was discussed previously in the interpretations of Figures 2, 3, and 4. The parameter estimates for absolute yard line are negative for field goals and positive for punts across both wins and losses. This means that as the absolute yard line decreases the likelihood of a field goal increases with respect to first down attempts, and as the absolute yard line increases the likelihood of a punt also increases. Yards to go has positive estimates for field goals and punts across both wins and losses. This also makes sense; as a positive relationship would indicate that as the yards to go increases, the likelihood that a team will choose to punt or kick a field goal also increases with respect to the likelihood that they will attempt a first down.

Minutes remaining and score margin are where we see the most obvious differences between the parameter estimates across wins and losses. The estimates for minutes remaining in losses are approximately two times the value of their estimates in wins, and the parameter estimates for score margin are 10 times the value for losses compared to wins. These are again logical results as they indicate that losing teams are less likely to punt or kick a field goal as their score margin and the time remaining in the game decrease. This makes sense because when teams are trailing by a large amount and/or time is running out, they have more of a necessity to score points quickly and thus, have less incentive to punt or kick a field goal.

It is also worth noting that the absolute yard line parameter estimates are very similar across wins and losses for both field goals (wins: -2.0734, losses: -2.0620) and punts (wins: 3.4359, losses:

3.4623), and four of the smallest estimates are the minutes remaining and score margin for punts and field goals within the winning teams' decisions. That information coupled with the information in the previous paragraph further indicates that teams are primarily choosing to attempt a first down out of desperation, rather than appropriately weighting all of the relevant information in their decision-making process.

5.3 Hypothesis Test

The hypothesis test presented in Equation (8) returned statistically significant results indicating to reject the null hypothesis. The parameter estimates for the original and new information were 0.1334 and 0.9693 respectively. The original information had a standard error of 0.025 and a t-statistics of 5.276 while the new information had a standard error of 0.15 and a t-statistic of 63.958 and both estimates have p-values less than 0.001, which indicates that these results are statistically significant. The parameter estimate for the constant was -0.1033 with a standard error of 0.017, and also had a p-value less than 0.001, which means that all three estimates for the hypothesis test are considered to be statistically significant.

These parameter estimates show that coaches are putting approximately seven times more weight on new information relative to original information, which signifies the presence of the representativeness heuristic in their decision-making processes on fourth down.

The same test for representativeness was conducted individually for each of the following subsets of data:

- the years contained in the validation set (even-numbered years)
- home and away teams separately
- each of the four quarters
- all 32 teams in the National Football League

Testing for representativeness within each season was done to determine whether or not the coaches were getting any better or worse at appropriately weighting the information from year to year. The results for each year can be seen in Table 6. Every year had a higher parameter estimate for the new information compared to the original information, which indicates the presence of representativeness in the coaches' decision-making processes for each year. In 2004 the difference between the new and original information parameter estimates was 0.8653, while in 2014 it was 0.8367, which indicates that the coaches, as a whole, have not improved their weighting of the two components in their decision-making processes on fourth down over the given timespan. The p-values for the new information were less than 0.001 for each year, indicating statistical significance, and all but one of the parameter estimates for the original information were significant at the 0.1 level, which is also true for the parameter estimates of the constant.

Comparing home and away teams with respect to the representativeness heuristic was done to see if coaches tend to be more or less susceptible to the bias when they are playing in their own stadium. These results can be seen in Table 7. For home games the parameter estimates for the new and original information are 0.9534 and 0.1584, respectively. Similarly, for away games the parameter estimates are 0.9569 for the new information and 0.1494 for the original information, while the estimates for the constant were -0.1088 for home teams and -0.0975 for away teams, and the p-values for all six estimates are less than 0.001, indicating statistical significance. With the parameter estimates being so similar across home teams and away teams, it appears that coaches, as a whole, do not weight the information differently when they are on the road compared to when they are at home.

Evaluating the individual quarters was done to see if coaches weighted information differently throughout the course of a game. The results for each quarter are shown in Table 8. For every quarter the new information has a higher coefficient estimate than the original information, indicating that the representativeness heuristic is present in the coaches' decision-making processes on fourth down regardless of when the fourth down is occurring during the game. The estimates of the weights are closest during the second and third quarters where they are different by approximately 0.53 for both quarters, while the first and fourth quarters see larger differences at 0.81 and 0.90, respectively.

The results were statistically significant at the 0.05 level for 10 of the 12 parameter estimates, and at the 0.01 level for eight of the 12 estimates. Original information in the fourth quarter has a p-value of 0.011, while original information in the first quarter had a p-value of 0.185, which could be due to the disproportionate amount of times coaches have chosen to kick in the first quarter rather than attempt to gain a first down, as discussed previously.

Each of the 32 teams in the NFL were tested separately to examine which teams were overweighting new information the most, and which were closest to weighting the new and original information optimally. The results for each team can be seen in Table 9. Though all 32 teams have a higher parameter estimate for the new information when compared to original information, the New Orleans Saints are the team with the smallest difference (0.1768) while the New York Jets had a difference of 1.6266, which is the largest gap out of all the teams. The p-values for the new information were less than 0.001 for every team, which indicates statistical significance. The original information had statistical significance at the 0.1 level for 12 parameter estimates. However, 15 had p-values of 0.3 or greater. While this means that those parameter estimates are not statistically significant, it does indicate that the correlation for those teams could be due to randomness, which could mean they are not taking the original information into account at all. Another point to support this thought is that the five teams with the lowest p-values for the original information are five of the most successful teams during the 12-year span covered in this dataset.

The Pittsburgh Steelers, New York Giants, and New England Patriots are the only teams to win multiple Super Bowls during the 12 years covered in this dataset, and all three of those teams are in the top four in terms of smallest difference between the parameter estimates for the new information and original information. The Saints also won a Super Bowl during this span, as did the Green Bay Packers (the sixth-best team on the list), meaning that five of the top six teams accounted for nine of the 12 Super Bowl victories in this timespan, and those are the same five teams mentioned above for having the lowest p-values for the original information. Conversely, the bottom five teams (the Jets along with the Arizona Cardinals, Minnesota Vikings, Chicago Bears, and Houston Texans) only made the playoffs an average of 3.2 times in the specified 12 years, with none of them making the playoffs more than four times. There are 12 playoff spots per year so on average teams made the playoffs 4.5 times during this stretch. The top six teams averaged seven playoff appearances.

It important to note that each of the equations presented in this research were conducted several different times while using different portions of the data as training sets, test sets, and validation sets. Sometimes different variables were included (or not included), or different transformations of those variables were used. This was done in an attempt to identify the most accurate models, capture all of the relevant effects, and as a result, generate the best representation of reality.

An example of this is shown in Table 10, which presents the results from Equation (9) across different variations of the in-game win odds model (Equation (4)). To test the reliability of the in-game win odds, the 12-year data set was separated into two subsets of six years each; one to build the model (a training set) and one to predict on (a validation set). This was done using four different groupings of six years: the first six years (2003-2008), last six years (2009-2014), odd-numbered years (2003, 2005, 2007, etc.), and even-numbered years (2004, 2006, 2008, etc.). All four variations yielded similar results in Equation (9), with the parameter estimates ranging from 5.0310 to 5.2605, the standard errors being either 0.014 or 0.015, and p-values of less than 0.001 for each.

Table 10 also shows the difference between generating the in-game win odds with respect to the home team compared to generating them with respect to the offensive team. Again, the results from Equation (9) were quite similar. The standard error was the same for both at 0.014, as was the p-value (0.000), while the parameter estimates are 5.1316 and 5.1861, when evaluating the home team and offensive team respectively. These comparisons were made for the inclusion and exclusion of different variables as well.

With similar results across the board for the different variations of the in-game win odds, the decision of which version to use for the final analysis was made using the logical judgment of which model variation was believed to be the best reflection of reality and minimized the introduction of randomness. The choice to use the odd-numbered years for the training set was made to help negate any underlying effects that could have directly or indirectly impacted the decision-making processes of coaches over the years, such as rules changes and the improved quality of kickers. For example, between the 2008 and 2009 seasons there were four new rules implemented to improve player safety (in addition to new rules regarding the replay system, field obstructions, and other circumstantial issues), and if any of those rules had an unforeseen or unrecognized impact on the likelihood of a team winning the game across different game-states, then that information would only be included in the validation set, but not in the test set, if the data were split into the first and last six years. Instead, using odd-numbered and even-numbered years eliminates the possibility of an issue of this nature occurring.

Each of the models developed in this research represented a different portion of information needed to make it possible to directly test for the representativeness heuristics. Each model returned statistically significant information from large data sets, which enabled the maximization of the validity of each component. Trying multiple variations of each model along the way and identifying the best version for generating the necessary information within each step of the process made it possible to bring them all together while minimizing the potential variation and error of the overall process.

There are a some specific examples that serve as interesting cases when examining the decisions that were actually made by coaches in real-game situations. The two that I will discuss are:

- the New Orleans Saints (hosting the Green Bay Packers in 2014)
- the New York Giants (hosting the Indianapolis Colts in 2006)

Both of the teams had to make a fourth-down decision three times in their specified game. The two examples listed are games in which the coach made the optimal choice for each of those three fourth down occurrences. These examples are important because they provide a lens through which we can examine the coaches when they are making optimal decisions on every fourth down they face over the course of a game.

When the New Orleans Saints played the Green Pay Packers in 2014 they never punted the ball. Of their three fourth downs, they kicked two field goals and had one first down attempt. The two field goals were both in the first half of the game while they were trailing by three points and they tied the game each time. The first down attempt was early in the third quarter when the game was tied. They were on the Packers' 43-yard line with two yards to go. They only gained one yard and therefore failed their first down attempt.

Despite the failed attempt, the Saints went on to win the game. This is an important example because it illustrates a situation in which the coach, Sean Payton, made the optimal decision, regardless of the attempt being unsuccessful. If a decision does not lead to the desired outcome that does not mean it was a poor or suboptimal decision. This is why decisions should be evaluated based on the decision-making process rather than evaluated on the outcome of that decision.

In the second game listed above, the Giants had one first down attempt, one field goal attempt, and one punt. They faced their first fourth down toward the end of the first quarter when they were trailing by three points and had the ball at the Colts' 33-yard line with five yards to go. The optimal decision was to attempt a first down, which is what their head coach, Tom Coughlin, opted to do. About midway through the second quarter they were down by six at the Colts' 22-yard line. The optimal choice was to kick a field goal, which is what they did. Lastly, trailing by two points late in the third quarter they had a fourth down at their own 31-yard line and they chose to punt, which was again the optimal decision.

Unfortunately, the Giants did not go on to win this game. They failed their first down attempts, missed their field goal, and after their punt they had turnovers on two of their next three possessions (a lost fumble and an interception), both of which resulted in the Colts scoring (a touchdown and a field goal) on their consequential next possessions. This again demonstrates how somebody can make an optimal decision without having the desired outcome. Even though they had several good decisions with poor outcomes, the Giants lost this game by just five points.

The obvious difference between these two games is that one resulted in a win while the other resulted in a loss. However, in games where a team made the optimal decision on more than half of their fourth downs, they won the game approximately two-thirds of the time (67.2 percent) and even teams that made the optimal decision between 25 and 50 percent of the time still won more often than not, winning 51.5 percent of their games. While the examples above used teams that only faced three fourth down decisions in those games, in this data set one team had 16 fourth downs in a single game, and on average teams faced more than seven (7.42) fourth downs per game.

Each decision can be extremely important. Improving from making two optimal decisions in a game to three resulted in the team's win percentage increase from 38.5 to 48.8 (10.3 percentage points), and going from three to four saw another large jump to 56.2 percent, which is an increase of 7.4 percentage points. A similar trend is observed when evaluating the *percentage* of optimal decisions a team makes in a game, rather than the *number* of optimal decisions they make in a game. Figure 6 illustrates a team's win percentage across their percentage of optimal decisions in a game. This is important because teams will face a varying number of fourth downs from game to game, and without knowing how many there will be in any given game, they should be striving to make the optimal decision as frequently as possible.

6 Discussion

The above results indicate that coaches are yielding to the representativeness heuristic when they are making fourth down decisions. This information could be useful for the coaches themselves as well as other high-level decision-makers within the organization. However, beyond just the context of sport, the methodology used in this research and the approach to directly test for the representativeness heuristic will be useful to future researchers in the fields of economics, psychology, politics, human resources and management, among others.

While previous empirical researchers were unable to determine the weight people were applying to new and prior information when examining real-world scenarios, this study keeps the two components of representativeness separate by utilizing a Bayesian approach, which in turn, allows us to directly test for the representativeness heuristic. This methodology introduces a new way for academics to research the representativeness heuristic without being confined to a laboratory setting (Ashton and Ashton, 1998; Charness and Levin, 2005; Charness et al., 2007; Friedman, 1998; Grether, 1980, 1992).

The results showing that coaches make suboptimal decisions on fourth down by acting too conservatively align with the findings of previous researchers (Carroll et al., 1989; Carter and Machol, 1971, 1978; Patel, 2012; Romer, 2002, 2006). Despite the fact that the majority of those researchers used an expected points model (Carroll et al., 1989; Carter and Machol, 1971, 1978; Romer, 2002, 2006) while only Patel (2012) used a win probability in addition to an expected points model, they all came to the same conclusion as this research in terms of determining if coaches are making suboptimal decisions on fourth down, even though they were unable to determine why coaches are making suboptimal decisions on fourth down.

It is remarkable that research on this topic has been going on since 1971, striving to statistically illustrates the idea that coaches in the National Football League are failing to regularly make the optimal decision on fourth down. It is even more perplexing when considering the fact that Virgil Carter (one of the authors of that 1971 article) played quarterback in the NFL from 1968 through 1976, meaning he was an active player on an NFL roster at the time his article was published. Despite this fact, coaches have continued to make suboptimal decisions on fourth down. Over the next 45 years researchers continued to demonstrate this (Carroll et al., 1989; Carter and Machol, 1978; Patel, 2012; Romer, 2002, 2006) without seeing a noteworthy difference in the coaches' behaviors.

This could be due to coaches not being made aware of the research that has been conducted or not understanding the biases and heuristics that affect the way our brains subconsciously process information. With this research being able to identify one of the factors impacting the coaches' decision-making processes, it may enable them to better recognize moments in which this bias is occurring. Fundamentally, as there continues to be more of a culture shift toward the acceptance and implementation of statistical analysis within the sport of football, and within the industry of sport as a whole, it could lead to more coaches having access to this kind of information, and in turn, lead to them making optimal decisions more often.

As previously mentioned, the results from Equation (1) and the subsequent recommendations for the optimal decision (shown in Figures 2, 3, and 4) strongly contradict conventional wisdom. It would therefore be unexpected for a coach to see these findings and begin to make fourth down decisions based solely on what was shown in this research. However, if the coaches were able to have meet regularly with an analytics specialist then they could gain a better understanding of the representativeness heuristic (as well as other subconscious psychological biases) and its role in how our brains process information. That would, in turn, help those coaches make optimal decisions more often, by enabling them to be more aware of how representativeness is impacting their decision-making and to better recognizing when they are in a situation in which this is occurring.

If a franchise in the NFL is willing to not only hire an analytics specialist, but also be committed to developing an environment conducive to integrating an analytics specialist into their organization's overall infrastructure, then that franchise would be best-suited to incorporate this research, and similar research, into their day-to-day operations. While this research alone would aid a team's in-game decision-making, creating that type of overall culture would also provide scouts with additional tools for player evaluation; help general managers prepare for the upcoming draft each year, compare the different ideas they have on roster composition, and assess the coaches themselves; and it would also assist the coaches when they develop their weekly strategies and gameplans as they prepare for each opponent throughout the season.

Furthermore, the coaches and the analytics specialist should be able to identify circumstances in which the coach is comfortable with implementing this research and the resulting recommendations for the optimal decision on fourth down. They should be able to have constructive conversations to gain a better understanding of each person's perspective on the matter, and thus, be able to find a middle ground that both parties are comfortable with and then work together to develop an in-game chart that the coaches would be able to quickly and easily reference when it comes time for them to actually make those fourth down decisions during a game.

This give-and-take dyadic relationship could and should be implemented across a wide array of potential topics, opening the door to a productive dialogue leading them to further identify situations in which the coach would like to have more information that the analytics specialist might be unaware of, as well as identifying the statistical tools and capabilities that the analytics specialist possesses that the coach might be unaware of. Creating an environment of this nature would allow the coaches and analytics specialist to develop a rapport and a level of trust that would help maximize the potential of implementing analytics into a NFL organization.

This research aligns with and reinforces the existing theory and the findings of previous researchers who studied the impact of the representativeness heuristic on people's decision-making processes. The results from the hypothesis test (Equation (8)) show that coaches are underweighting the prior information (Grether, 1980, 1992), and that they are not efficiently incorporating the new information (Charness and Levin, 2005; Friedman, 1998) when making these fourth down decisions.

Within the broader context of Prospect Theory, this research shows how challenging it would be for someone to quickly estimate these kinds of values with acceptable accuracy. This further illustrates one of the underlying concepts of Prospect Theory, which is the idea that bias is a byproduct of the fact that it is too difficult and computationally intensive for people to always attempt to maximize utility (Kahneman and Tversky, 1979; Tversky and Kahneman, 1973, 1974).

Utilizing a Bayesian approach for not only the decision model, but for the win probability model as well, could help answer other questions or complications that have come up in previous research. Bayes' rule includes a prior likelihood, which in this case allows us to account for team quality when estimating in-game win probabilities. Including a prior also helps regulate any extreme estimations due to limited data within specific situations. Being able to account for this situational data at all while analyzing decision-making is another benefit of a Bayesian approach.

Previous researchers (Carroll et al., 1989; Carter and Machol, 1971, 1978; Patel, 2012; Romer, 2002,

2006) avoided accounting for such information in their expected points models by examining only plays that took place in a certain quarter of the game, within a certain score margin, or between certain yard lines, and none of them accounted for team quality. Future researchers would be able to include more in-game situations (and thus, more data) in their analyses by using a Bayesian approach, whether it be to develop a win probability model, an expected points model, or another means by which to quantify the value of a given decision or situation.

Beyond NFL coaches' decision-making, this research and its results are potentially generalizable to the influence of the representativeness heuristic on peoples' decision-making under uncertainty. NFL coaches are highly qualified personnel and among the best in their field. Coaches and their decisions are highly scrutinized.

The accuracy of the coaches' decisions is critiqued by the fans, the media, the players, and even other coaches. Franchises in the NFL regularly dismiss head coaches who are failing to meet the expectations that the organization is holding them to. If other NFL teams consider those coaches to still be capable of living up to their standards then one of those teams might choose to hire him. Otherwise, that coach is left to retire or seek employment elsewhere that would be of a lesser role, at a lower level of competition, or within a different sector of the industry entirely.

Conversely, the teams that are looking to replace their coach will look to hire someone who has been impressive in a lesser role, successful at a lower level of competition, or has proven themselves a worthy candidate as the head coach for another team in the NFL. Overall, this process leads to an efficient system for ensuring that the 32 head coaches in the NFL are among the best football coaches in the world.

This research aimed to analyze the decisions made by NFL coaches and the decision-making processes that led them to make those choices. This means that the immediate outcome of a play was not included in the analysis, but only the eventual outcome of the game. This study examined how the decision itself would impact a team's probability of winning rather than how the outcome of a play would impact the team's probability of winning. This is in contrast to the other researchers (Carroll et al., 1989; Carter and Machol, 1971, 1978; Patel, 2012; Romer, 2002, 2006), who made assumptions about the situation a team would be in, given that a play was successful or unsuccessful, and then, based on those assumed situations, estimated the costs and benefits for each of the potential decisions.

Therefore, it is somewhat unexpected (but encouraging) that the findings from this research and the resulting estimations for when the option to attempt a first down becomes the optimal decision would so closely reflect theirs. It is important to note that, despite the numerous differences in our methodologies, our results are still similar. It tells us that no matter how one dissects the question, coaches are making suboptimal decisions, and no matter the methodology for determining what the optimal decision is, the resulting recommendations are similar.

However, where the other researchers (Carroll et al., 1989; Carter and Machol, 1971, 1978; Patel, 2012; Romer, 2002, 2006) had to make inferences based on the contexts of their findings, this analysis directly tests for a specific subconscious psychological bias. Due to their methodologies, the other researchers were unable to determine why the actual decisions being made by coaches did not reflect those of the aforementioned recommendations, and therefore their results were open to criticism.

Alternatively, previous researchers examining the representativeness heuristic (Ashton and Ashton 1998; Charness and Levin, 2005; Charness et al., 2007; Friedman, 1998; Grether, 1980, 1992) conducted laboratory experiments to directly test for the presence of a bias (or biases) in their subjects' decision-making processes. Though previous researchers had to pick between studying a real-world scenario and being able to directly test for a subconscious psychological bias, this research utilized observational data from a real-world scenario while retaining the ability to directly test for the representativeness heuristic.

7 Concluding Remarks

Limitations for this study include the idea that when teams choose to attempt a first down on fourth down they are already in a situation where they are likely to lose the game regardless of whether the decision being made is the optimal choice or not. This could potentially be addressed with future research that would use a Bayesian approach to create an expected points model rather than a win probability model.

Additionally, it may have been better to use a Bayesian regression for the conditional likelihood components rather than multinomial logistic regressions. This is also something that could be taken into account for future research. It could also be beneficial to look at specific coaches or coaching characteristics such as age and experience, rather than pooling all of the coaches together and analyzing them as one group.

Another area for future research could address other potential factors that might impact a coach's decision-making process, such as social pressures, in-game injuries, season-level objective functions, and coach-quarterback interaction. These ideas could lend themselves to future research such as case studies, examinations of the principal-agent concept, and several other possibilities.

Future research could also include testing for the presence of additional heuristics and psychological biases in the coaches' decision-making processes beyond the representativeness heuristic alone. This research has motivated another project to test for loss aversion, which is currently being worked on.

This research utilized an innovative way to directly test for the representativeness heuristic in real-world decision-making. The methodology used for this paper could impact not only sport analytics, but a variety of other fields as well, including economics, psychology, politics, human resources, and management among others.

While the method and results from this research will be of interest to sport managers and analytic practitioners, the primary contribution is in the discernment of optimal decision-making relating to outcomes. It is certain that sport psychology will be interested in the theory, method, and results; as will organizational theorists concerned with investigating decisions within the greater context of organizations. This research also helps to further demonstrate that in-game sport contest information is a fruitful area for academics to examine and test hypotheses.

The results of this research are potentially generalizable beyond NFL coaches' decision-making processes. This is, in part, because these findings coincide with the claims of the previous researchers who examined the influence of the representative heuristic on how people process information and the decisions that they make under uncertainty. Additionally, this provides further support against the criticisms of those previous researchers' conclusions.

Within the context of this paper, there are implications for the decision-making processes of coaches, who could be more conscientious about how they are allowing different information to impact the choices they make. Simply being made aware of the existence of the representativeness heuristic and other psychological biases could help coaches make optimal decisions more often.

Beyond the coaches themselves, this research would be beneficial to general managers and other high-level decision-makers within NFL front offices, who could use this information to aid in their evaluations of coaches. This could be utilized when determining whether or not to retain the team's current head coach and when choosing between a pool of candidates to hire as a new head coach.

Additionally, the Bayesian approach used for this research will be useful to future academics who wish to directly test for the representativeness heuristic. The hope is that the methodology utilized for this research will further progress the research that is being done not only within the field of sport analytics, but within the realm of sport management as a whole.

8 Graphs and Tables

Table 1. Summary statistics of the game-state information across the three options a team has on fourth down: first down attempts, field goal attempts, and punts.

Game-state component	Summary data				
	n	Mean	Std. dev.	Min.	Max.
First Down Attempt					
Minutes remaining	5790	18.0888	17.5321	0.0170	59.7330
Absolute yard line	5790	35.9477	21.7339	1.0000	99.0000
In game win prob	5790	0.3416	0.3306	0.0004	0.9996
Score margin	5790	-6.5568	13.1643	-59.0000	59.0000
Pre-game win prob	5790	0.4680	0.2022	0.0207	0.9793
Point Total	5790	42.7721	4.7605	30.0000	60.0000
Defense Timeouts	5790	2.3775	0.8601	0.0000	3.0000
Offense Timeouts	5790	2.1385	1.0000	0.0000	3.0000
Yards to go	5790	4.7230	4.9080	1.0000	35.0000
Home Indicator	5790	0.4769	0.4995	0.0000	1.0000
Field Goal Attempts					
Minutes remaining	10583	28.9001	16.3236	0.0000	59.1170
Absolute yard line	10583	18.2816	9.9141	1.0000	46.0000
In game win prob	10583	0.5440	0.2945	0.0005	0.9993
Score margin	10583	0.7259	9.4222	-45.0000	48.0000
Pre-game win prob	10583	0.5109	0.2000	0.0207	0.9724
Point Total	10583	42.8035	4.7670	30.0000	60.0000
Defense Timeouts	10583	2.5449	0.7859	0.0000	3.0000
Offense Timeouts	10583	2.4372	0.8265	0.0000	3.0000
Yards to go	10583	6.8757	4.6140	1.0000	34.0000
Home Indicator	10583	0.5134	0.4998	0.0000	1.0000
Punts					
Minutes remaining	29186	30.5721	16.9582	0.0500	59.9000
Absolute yard line	29186	64.8480	14.8729	27.0000	99.0000
In game win prob	29186	0.5051	0.3071	0.0006	0.9997
Score margin	29186	-0.3925	10.5610	-59.0000	59.0000
Pre-game win prob	29186	0.4856	0.2020	0.0207	0.9793
Point Total	29186	42.4049	4.7405	30.0000	60.0000
Defense Timeouts	29186	2.5961	0.7516	0.0000	3.0000
Offense Timeouts	29186	2.6494	0.6484	0.0000	3.0000
Yards to go	29186	8.5331	5.9038	1.0000	48.0000
Home Indicator	29186	0.4857	0.4998	0.0000	1.0000

Table 2. Summary statistics for pre-game win probabilities based on game outcomes, presenting the mean, standard deviation, minimum, and maximum across wins, losses and ties.

Outcome	n	Mean	Std. Dev.	Min.	Max.
Win	1760	0.6356	0.1695	0.1121	0.9793
Loss	1308	0.4891	0.1825	0.0613	0.8979
Tie	4	0.6234	0.2876	0.2078	0.8677

Table 3. Summary statistics of the game-state data used to estimate the conditional likelihood across wins, losses and ties, with respect to the home team when they are on offense.

Component	Outcome	Summary data				
		n	Mean	Std. Dev.	Min.	Max.
		Offense				
Down	Loss	73017	2.0203	1.0191	1.000	4.000
	Tie	244	1.9303	1.0099	1.000	4.000
	Win	104715	1.9776	0.9955	1.000	4.000
Absolute yard line	Loss	73017	53.4923	23.7054	1.000	99.000
	Tie	244	51.8689	22.7167	1.000	92.000
	Win	104715	48.9775	24.6831	1.000	99.000
Minutes remaining	Loss	73017	28.6493	17.5820	0.000	60.000
	Tie	244	29.3236	17.6829	0.033	59.917
	Win	104715	29.5292	17.2416	0.000	60.000
Point Total	Loss	73014	42.7897	4.7989	31.000	57.000
	Tie	244	42.0820	2.5112	38.000	44.500
	Win	104711	42.8366	4.7809	30.000	60.000
Defense Timeouts	Loss	73017	2.5811	0.7040	0.000	3.000
	Tie	244	2.7295	0.5598	0.000	3.000
	Win	104715	2.5824	0.7390	0.000	3.000
Offense Timeouts	Loss	73017	2.5689	0.7826	0.000	3.000
	Tie	244	2.6926	0.7697	0.000	3.000
	Win	104715	2.6175	0.6882	0.000	3.000
Yards to go	Loss	73017	8.6102	4.0640	1.000	43.000
	Tie	244	8.7664	3.5623	1.000	23.000
	Win	104715	8.3077	3.9246	1.000	46.000
Score margin	Loss	73017	-6.4405	8.8818	-46.000	24.000
	Tie	244	-2.8648	6.0029	-16.000	7.000
	Win	104715	5.3618	9.2666	-24.000	59.000

Table 4. Summary statistics of the game-state data used to estimate the conditional likelihood across wins, losses and ties, with respect to the home team when they are on defense.

Component	Outcome	Summary data				
		n	Mean	Std. Dev.	Min.	Max.
		Defense				
Down	Loss	69218	2.0168	1.0045	1.000	4.00
	Tie	237	2.0253	1.0165	1.000	4.00
	Win	93629	2.0558	1.0226	1.000	4.00
Absolute yard line	Loss	69218	50.6768	24.0794	1.000	99.00
	Tie	237	48.6920	23.9215	2.000	99.00
	Win	93629	46.5516	23.1833	1.000	99.00
Minutes remaining	Loss	69218	29.2596	17.2849	0.017	60.00
	Tie	237	27.1716	17.6969	0.100	59.25
	Win	93629	28.7268	17.5063	0.017	60.00
Point Total	Loss	69218	42.8003	4.8311	31.000	57.00
	Tie	237	41.9177	2.5505	38.000	44.50
	Win	93629	42.8526	4.7845	30.000	60.00
Defense Timeouts	Loss	69218	2.6176	0.7217	0.000	3.00
	Tie	237	2.6920	0.6843	0.000	3.00
	Win	93629	2.6317	0.6640	0.000	3.00
Offense Timeouts	Loss	69218	2.5711	0.7231	0.000	3.00
	Tie	237	2.6287	0.7287	0.000	3.00
	Win	93629	2.5390	0.7896	0.000	3.00
Yards to go	Loss	69218	8.2970	3.9963	1.000	41.00
	Tie	237	8.2321	3.8993	1.000	23.00
	Win	93629	8.5866	4.1736	1.000	48.00
Score margin	Loss	69218	-3.9189	8.7525	-46.000	25.00
	Tie	237	-0.4219	5.1765	-13.000	10.00
	Win	93629	7.8970	9.4279	-24.000	59.00

Table 5. Results of the multinomial logistic regression shown in Equation (5), which represents the conditional likelihood component of Equation (1), the decision model.

Component	Decision	Regression Results		
		Coefficient	S.E.	P-value
	Wins (n=10901)			
Constant	Field Goal	6.3035	0.339	0.000
	Punt	-11.5021	0.433	0.000
Minutes Remaining	Field Goal	1.1337	0.092	0.000
	Punt	1.4898	0.085	0.000
Score margin	Field Goal	0.0018	0.006	0.754
	Punt	0.0147	0.005	0.005
Absolute yard line	Field Goal	-2.0734	0.084	0.000
	Punt	3.4359	0.103	0.000
Yards to go	Field Goal	2.2989	0.069	0.000
	Punt	1.5007	0.055	0.000
Home indicator	Field Goal	-0.7343	0.106	0.000
	Punt	-0.2126	0.094	0.024
Offense Timeouts	Field Goal	-0.5792	0.079	0.000
	Punt	-0.3150	0.075	0.000
Defense Timeouts	Field Goal	-0.7277	0.086	0.000
	Punt	-1.1389	0.080	0.000
Offense Point Total	Field Goal	-0.0926	0.016	0.000
	Punt	-0.1479	0.014	0.000
Defense Point Total	Field Goal	-0.0334	0.018	0.062
	Punt	-0.0397	0.015	0.010
	Losses (n=12087)			
Constant	Field Goal	4.8133	0.314	0.000
	Punt	-14.9857	0.439	0.000
Minutes Remaining	Field Goal	2.9334	0.114	0.000
	Punt	2.3252	0.090	0.000
Score margin	Field Goal	0.1633	0.006	0.000
	Punt	0.1159	0.005	0.000
Absolute yard line	Field Goal	-2.0620	0.075	0.000
	Punt	3.4623	0.095	0.000
Yards to go	Field Goal	1.3213	0.057	0.000
	Punt	0.8614	0.043	0.000
Home indicator	Field Goal	-0.8468	0.096	0.000
	Punt	-0.2138	0.079	0.007
Offense Timeouts	Field Goal	-0.2372	0.062	0.000
	Punt	0.3221	0.055	0.000
Defense Timeouts	Field Goal	-1.1550	0.078	0.000
	Punt	-0.9966	0.069	0.000
Offense Point Total	Field Goal	-0.2879	0.019	0.000
	Punt	-0.2017	0.015	0.000
Defense Point Total	Field Goal	-0.0603	0.016	0.000
	Punt	-0.0206	0.014	0.140

Table 6. Results of the hypothesis test shown in Equation (8), separated by year to demonstrate the variation of the weight being applied to new and original information over time.

Year	Information	Regression Results		
		Coefficient	S.E.	P-value
2004	Constant	-0.1230	0.043	0.004
	New Information	0.9817	0.040	0.000
	Original Information	0.1164	0.068	0.089
2006	Constant	-0.0683	0.041	0.096
	New Information	0.8527	0.037	0.000
	Original Information	0.1588	0.065	0.015
2008	Constant	-0.1234	0.044	0.005
	New Information	1.0601	0.044	0.000
	Original Information	0.0527	0.073	0.470
2010	Constant	-0.0645	0.043	0.131
	New Information	0.9552	0.040	0.000
	Original Information	0.2867	0.067	0.000
2012	Constant	-0.1339	0.042	0.002
	New Information	0.9240	0.038	0.000
	Original Information	0.1633	0.067	0.014
2014	Constant	-0.1069	0.045	0.017
	New Information	0.9795	0.039	0.000
	Original Information	0.1428	0.065	0.028

Table 7. Results of the hypothesis test shown in Equation (8), separated by home and away teams, demonstrating the variation of the weight being applied to new and original information.

Team	1lInformation	Regression Results		
		Coefficient	S.E.	P-value
Home	Constant	-0.1088	0.026	0.000
	New Information	0.9534	0.023	0.000
	Original Information	0.1584	0.038	0.000
Away	Constant	-0.0975	0.024	0.000
	New Information	0.9569	0.023	0.000
	Original Information	0.1494	0.039	0.000

Table 8. Results of the hypothesis test shown in Equation (8), separated by quarter, demonstrating the variation of the weight being applied to new and original information.

Information	n	Regression Results			
		Coefficient	S.E.	t-value	P-value
		First Quarter			
Constant	5017	-0.1097	0.038	-2.882	0.004
New Information	5017	0.9250	0.043	21.291	0.000
Original Information	5017	0.1194	0.090	1.326	0.185
		Second Quarter			
Constant	6195	-0.0710	0.030	-2.350	0.019
New Information	6195	0.8348	0.035	23.653	0.000
Original Information	6195	0.3077	0.066	4.690	0.000
		Third Quarter			
Constant	5031	-0.1440	0.041	-3.529	0.000
New Information	5031	0.8601	0.042	20.703	0.000
Original Information	5031	0.3277	0.095	3.454	0.001
		Fourth Quarter			
Constant	6358	-0.0638	0.070	-0.906	0.365
New Information	6358	1.0585	0.030	34.879	0.000
Original Information	6358	0.1623	0.064	2.533	0.011

Table 9. Results of the hypothesis test shown in Equation (8), separated by team, to show the variation of the weights being applied to the information from team to team.

	Regression Results					
	New Information			Original Information		
Team	Coefficient	S.E.	P-value	Coefficient	S.E.	P-value
Arizona Cardinals	1.2444	0.102	0.000	-0.2529	0.155	0.103
Atlanta Falcons	1.3013	0.120	0.000	0.3579	0.164	0.029
Baltimore Ravens	0.9779	0.092	0.000	0.0944	0.158	0.550
Buffalo Bills	1.0635	0.098	0.000	0.1337	0.161	0.406
Carolina Panthers	0.9989	0.095	0.000	-0.0310	0.161	0.848
Chicago Bears	1.2303	0.113	0.000	-0.1413	0.168	0.401
Cincinnati Bengals	0.9615	0.089	0.000	0.3682	0.153	0.016
Cleveland Browns	0.8403	0.087	0.000	0.1709	0.152	0.259
Dallas Cowboys	0.7633	0.090	0.000	0.2874	0.151	0.056
Denver Broncos	0.9366	0.092	0.000	0.3907	0.163	0.017
Detroit Lions	0.8563	0.090	0.000	0.3358	0.160	0.036
Green Bay Packers	0.8955	0.090	0.000	0.3896	0.160	0.015
Houston Texans	1.2052	0.108	0.000	-0.0031	0.173	0.986
Indianapolis Colts	0.9371	0.106	0.000	-0.2202	0.203	0.277
Jacksonville Jaguars	0.9313	0.091	0.000	-0.0765	0.146	0.600
Kansas City Chiefs	0.9695	0.092	0.000	-0.0400	0.163	0.806
Miami Dolphins	0.9764	0.091	0.000	0.3169	0.155	0.041
Minnesota Vikings	1.4130	0.124	0.000	-0.0517	0.173	0.765
New England Patriots	0.9962	0.114	0.000	0.6833	0.186	0.000
New Orleans Saints	0.7357	0.084	0.000	0.5589	0.152	0.000
New York Giants	0.7079	0.079	0.000	0.4467	0.151	0.003
New York Jets	1.5622	0.139	0.000	-0.0644	0.183	0.725
Oakland Raiders	1.0265	0.102	0.000	0.2938	0.169	0.082
Philadelphia Eagles	0.8970	0.089	0.000	0.2301	0.161	0.152
Pittsburgh Steelers	0.6700	0.086	0.000	0.4391	0.155	0.005
San Diego Chargers	0.8672	0.092	0.000	0.1418	0.148	0.337
Seattle Seahawks	0.8145	0.079	0.000	0.1453	0.142	0.308
San Francisco 49ers	0.7415	0.073	0.000	0.0262	0.139	0.850
St. Louis Rams	0.8345	0.083	0.000	0.0314	0.152	0.837
Tampa Bay Buccaneers	0.9903	0.098	0.000	0.0156	0.155	0.920
Tennessee Titans	0.7742	0.078	0.000	0.1206	0.142	0.395
Washington Redskins	0.6966	0.077	0.000	0.1481	0.136	0.277

Table 10. Results of the logistic regression shown in Equation (9) for different specification of the in-game win odds model.

Variation	Regression Results		
	Coefficient	S.E.	P-value
Years in Training Set			
First six years	5.0310	0.014	0.000
Last six years	5.2605	0.015	0.000
Odd years	5.1316	0.014	0.000
Even years	5.1639	0.014	0.000
Team Evaluated in Training Set			
Home team	5.1861	0.014	0.000
Offensive team	5.1316	0.014	0.000

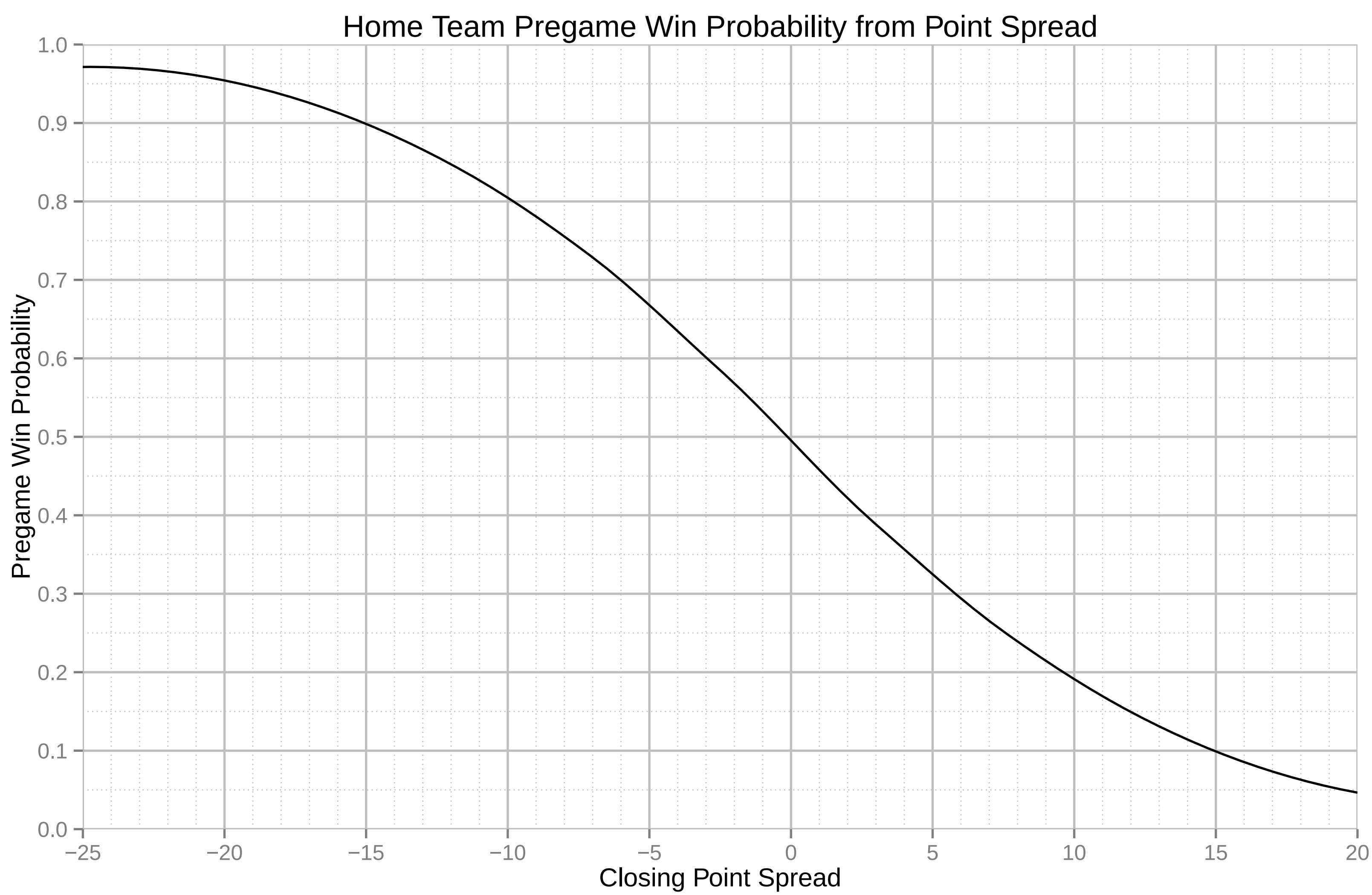

Home Team Pregame Win Probability from Point Spread
Pregame Win Probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−25
−20
−15
−10
−5
0
5
10
15
20
Closing Point Spread

Figure 1

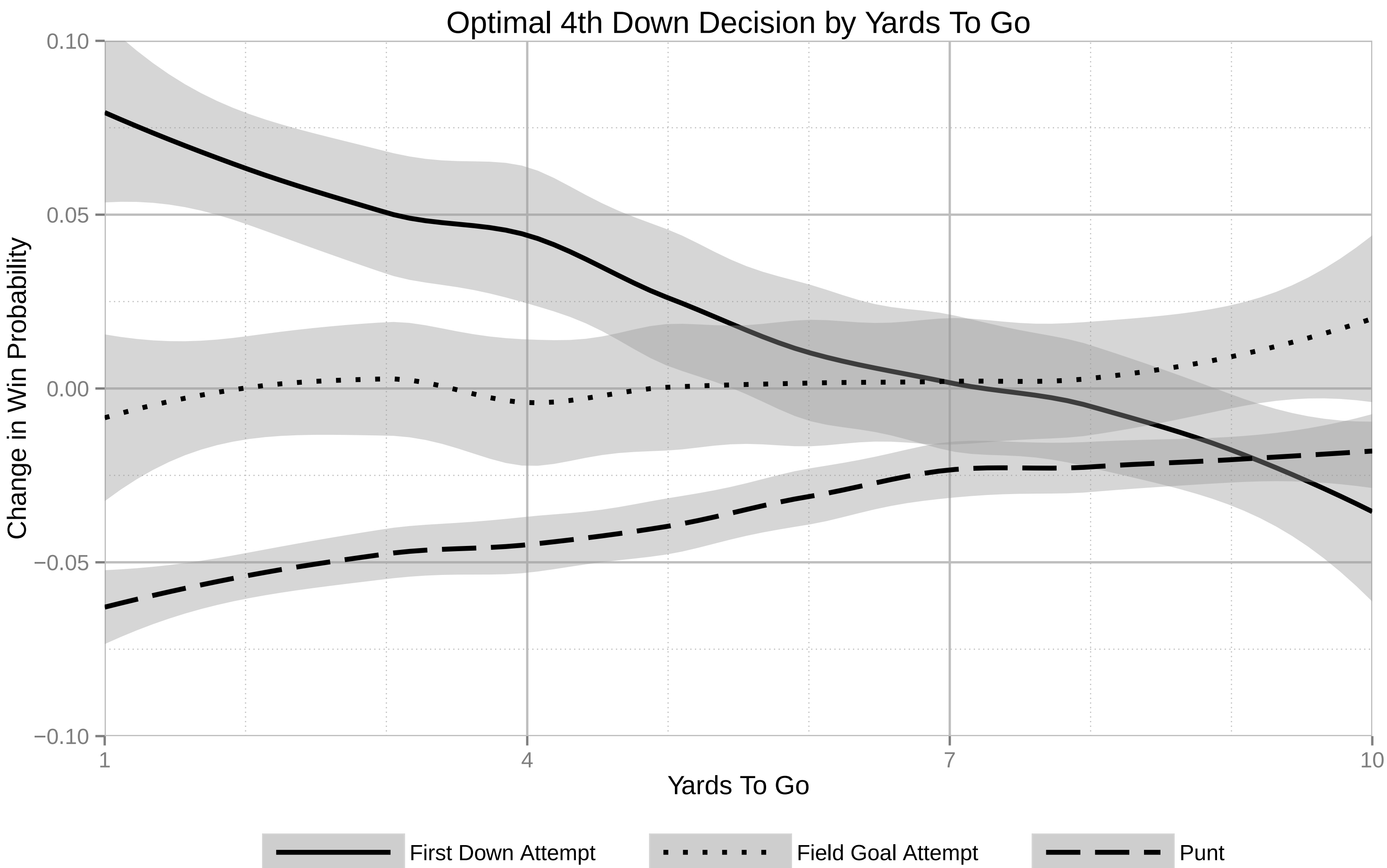
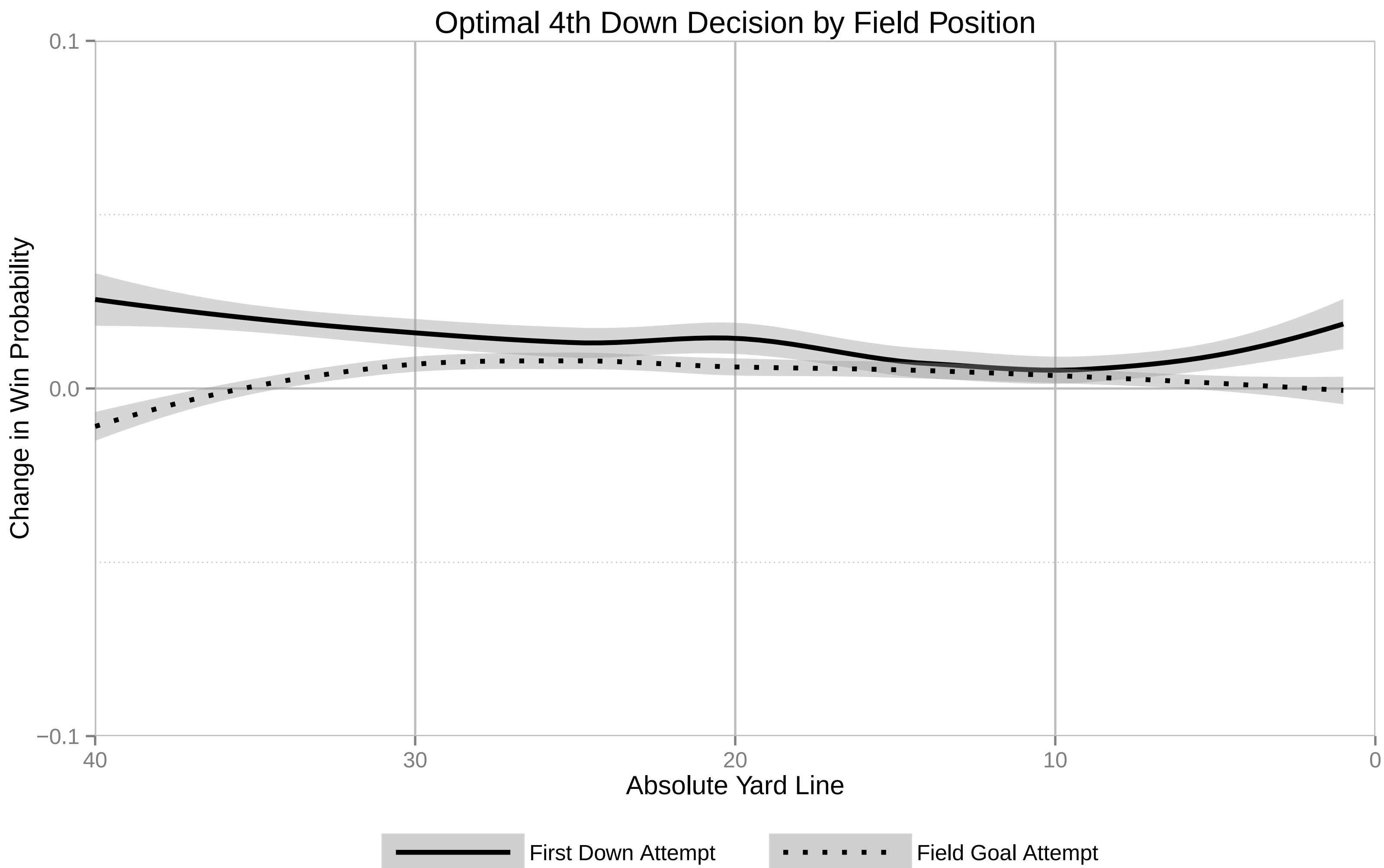

Optimal 4th Down Decision by Yards To Go
0.10
0.05
0.00
−0.05
−0.10
Change in Win Probability
1
4
7
10
Yards To Go
First Down Attempt
Field Goal Attempt
Punt

Figure 2



Optimal 4th Down Decision by Field Position
0.1
0.0
−0.1
Change in Win Probability
40
30
20
10
0
Absolute Yard Line
First Down Attempt
Field Goal Attempt

Figure 3

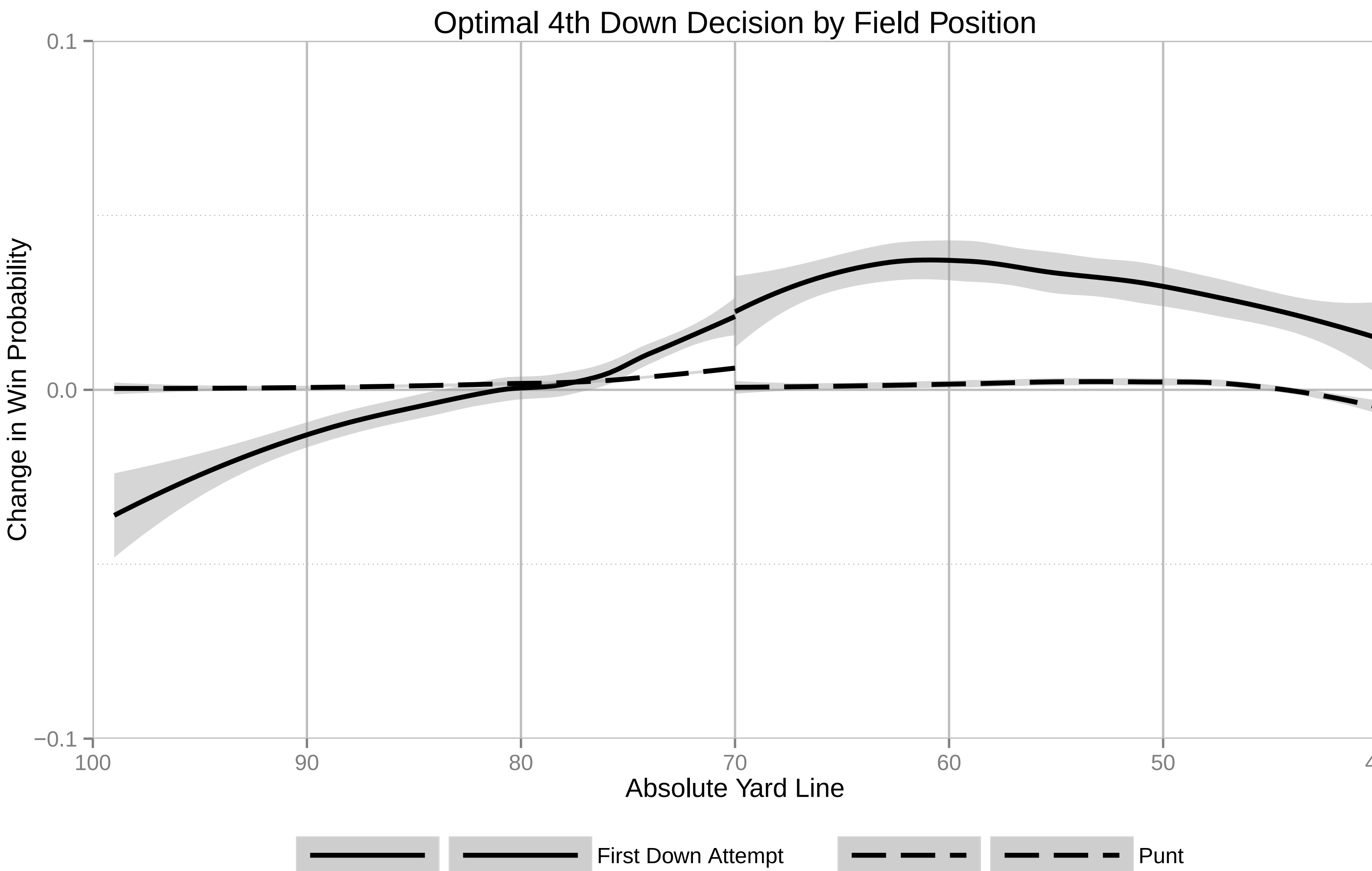

Optimal 4th Down Decision by Field Position
0.1
0.0
−0.1
Change in Win Probability
100
90
80
70
60
50
40
Absolute Yard Line
First Down Attempt
Punt

Figure 4

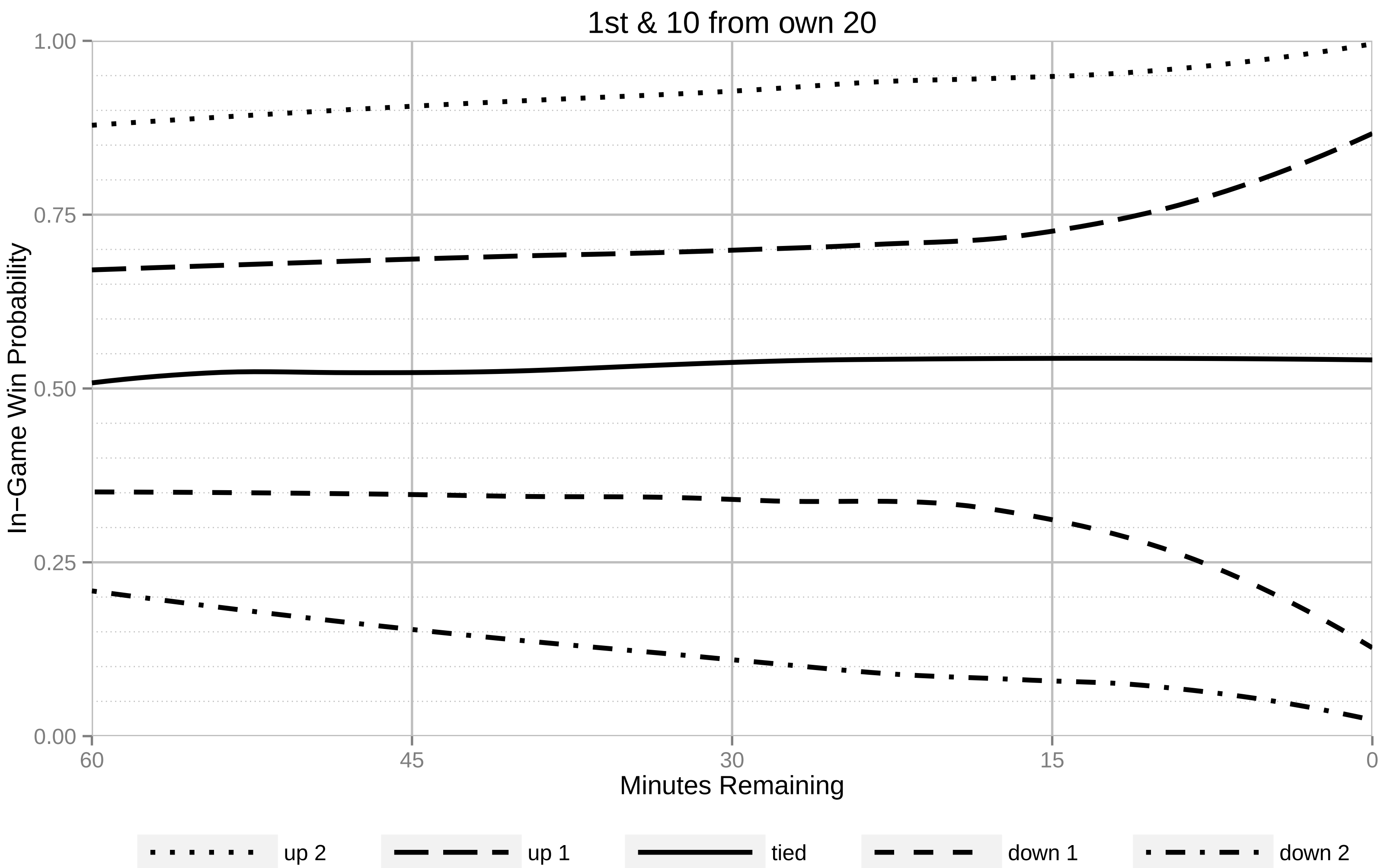

1st & 10 from own 20
In−Game Win Probability
1.00
0.75
0.50
0.25
0.00
60
45
30
15
0
Minutes Remaining
up 2
up 1
tied
down 1
down 2

Figure 5

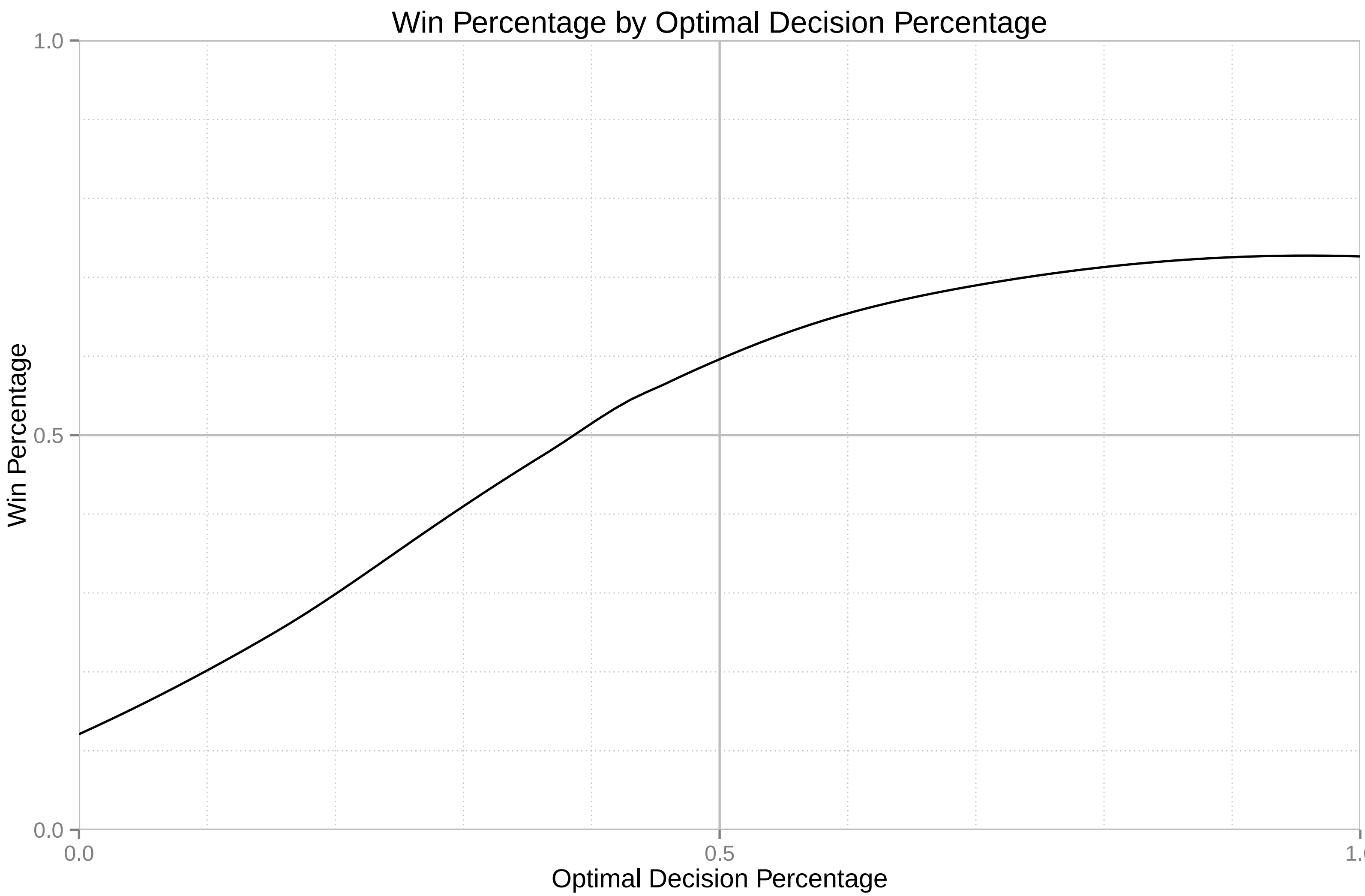

Win Percentage by Optimal Decision Percentage
1.0
0.5
0.0
Win Percentage
0.0
0.5
1.0
Optimal Decision Percentage

Figure 6

References

Alamar, B. (2013). *Sports analytics: A guide for coaches, managers, and other decision makers.* Columbia University Press.

Alamar, B. C. (2006). The passing premium puzzle. *Journal of Quantitative Analysis in Sports*, 2(4).

Alamar, B. C. (2010). Measuring risk in nfl playcalling. *Journal of Quantitative Analysis in Sports*, 6(2).

Ashton, A. H. and Ashton, R. H. (1988). Sequential belief revision in auditing. *Accounting Review*, pages 623–641.

Ashton, R. H. and Ashton, A. H. (1990). Evidence-responsiveness in professional judgment: Effects of positive versus negative evidence and presentation mode. *Organizational Behavior and Human Decision Processes*, 46(1):1–19.

Beasley, B., Greenwald, R., and Agha, N. (2015). Nfl time management: The role of timeouts in end-game scenarios. *The Journal of SPORT*, 4(1):49–64.

Bernoulli, D. (1954 [1738]). Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, pages 23–36.

Boronico, J. S. and Newbert, S. L. (1999). Play calling strategy in american football: A game-theoretic stochastic dynamic programming approach. *Journal of Sport Management*, 13(2).

Boronico, J. S. and Newbert, S. L. (2001). An empirically driven mathematical modelling analysis for play calling strategy in american football. *European Sport Management Quarterly*, 1(1):21–38.

Carroll, B., Palmer, P., and Thorn, J. (1989). *The hidden game of football.* Warner books.

Carter, V. and Machol, R. E. (1971). Technical note?operations research on football. *Operations Research*, 19(2):541–544.

Carter, V. and Machol, R. E. (1978). Note-optimal strategies on fourth down. *Management Science*, 24(16):1758–1762.

Charness, G., Karni, E., and Levin, D. (2007). Individual and group decision making under risk: An experimental study of bayesian updating and violations of first-order stochastic dominance. *Journal of Risk and uncertainty*, 35(2):129–148.

Charness, G. and Levin, D. (2005). When optimal choices feel wrong: A laboratory study of bayesian updating, complexity, and affect. *American Economic Review*, pages 1300–1309.

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

Friedman, D. (1998). Monty hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review*, pages 933–946.

Fry, M. J. and Ohlmann, J. W. (2012). Introduction to the special issue on analytics in sports, part i: General sports applications. *Interfaces*, 42(2):105–108.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, pages 537–557.

Grether, D. M. (1992). Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1):31–57.

Hadley, L., Poitras, M., Ruggiero, J., Knowles, S., et al. (2000). Performance evaluation of national football league teams. *Managerial and Decision Economics*, 21(2):63–70.

Jordan, J. D., Melouk, S. H., Perry, M. B., et al. (2009). Optimizing football game play calling. *Journal of Quantitative Analysis in Sports*, 5(2):1–34.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291.

Köbberling, V. and Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1):119–131.

Lopez, M. and Snyder, K. (2013). Biased impartiality among national hockey league referees. *Available at SSRN 2259798.*

Mongeon, K. (2015). A market test for ethnic discrimination in the national hockey league a game-level panel data approach. *Journal of Sports Economics*, 16(5):460–481.

Mongeon, K. and Longley, N. (2015). Testing for ethnicity discrimination among nhl referees: A duration model approach. *Eastern Economic Journal*, 41(1):86–101.

Mongeon, K. and Mittelhammer, R. (2011). Home advantage, close games, and big crowds: economics and existence of rationally biased officiating. Technical report, Working paper.

Mongeon, K. and Mittelhammer, R. (2013). To discriminate or not to discriminate–is data aggregation the question? *Applied Economics Letters*, 20(16):1485–1490.

Nash, J. (1951). Non-cooperative games. *Annals of mathematics*, pages 286–295.

Patel, S. (2012). *Do Coaches in the National Football League Exhibit Optimal Behaviour on Fourth Down Situations? An Economic Insight into NFL Decision-Making using the Expected Point and Win Probability Models of Utility*. PhD thesis, The University of Amsterdam.

Porter, R. C. (1967). Extra-point strategy in football. *The American Statistician*, 21(5):14–15.

Rockerbie, D. W. (2008). The passing premium puzzle revisited. *Journal of Quantitative Analysis in Sports*, 4(2).

Romer, D. (2002). It's fourth down and what does the bellman equation say? a dynamic programming analysis of football strategy. Technical report, National Bureau of Economic Research.

Romer, D. (2006). Do firms maximize? evidence from professional football. *Journal of Political Economy*, 114(2):340–365.

Sackrowitz, H. and Sackrowitz, D. (1996). Time management in sports: ball control and other myths. *Chance*, 9(1):41–49.

Sahi, S. and Shubik, M. (1988). A model of a sudden-death field-goal football game as a sequential duel. *Mathematical Social Sciences*, 15(3):205–215.

Stern, H. (1991). On the probability of winning a football game. *The American Statistician*, 45(3):179–183.

Sundali, J. and Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, 1(1):1.

Tom, S. M., Fox, C. R., Trepel, C., and Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, 315(5811):515–518.

Tversky, A. and Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive psychology*, 5(2):207–232.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323.

Urschel, J. D. and Zhuang, J. (2011). Are nfl coaches risk and loss averse? evidence from their use of kickoff strategies. *Journal of Quantitative Analysis in Sports*, 7(3).

Von Neumann, J. and Morgenstern, O. (1945). Theory of games and economic behavior. *Bull. Amer. Math. Soc*, 51(7):498–504.

Von Neumann, J. and Morgenstern, O. (2007). *Theory of games and economic behavior*. Princeton university press.

Winston, W. L. (2012). *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football*. Princeton University Press.