Optimal Lineups in Penalty Kick Shootouts: An Empirical and Theoretical Investigation

Lucas Hubbard and Christian Britto[1]

Professor Attila Ambrus, Faculty Advisor

[1] The authors can be contacted at lucashubbard1234@gmail.com and christian.britto@gmail.com, respectively. Starting in the fall, Mr. Hubbard will be working in Arlington, Va. at Applied Predictive Technologies as a Business Consultant, and Mr. Britto will work as a Venture For America Fellow at ChapterSpot in New Orleans.

Acknowledgements

Special thanks to Michelle Connolly and Kent Kimbrough for providing valuable feedback throughout the year, to Attila Ambrus for advising this work, and to NationalFootballTeams.com and the Rec Sport Soccer Statistics Foundation for their fastidious record-keeping. Finally, thanks to Portugal manager Paulo Bento for putting Cristiano Ronaldo last in his penalty shootout lineup in the 2012 Euro Cup Semifinals and providing the inspiration for this investigation.

Abstract

This paper shows how teams should order their lineups in a five-man penalty kick shootout. We begin with a theoretical investigation of how certain players' comparative advantage in stressful situations dictates optimal penalty lineup strategies for managers. We observe factors that affect a given player's success rate, including experience and the pressure of the situation. We recognize that more experienced shooters do not respond adversely to the pressure of a penalty shootout, while inexperienced players suffer a statistically significant adverse response. We conclude with a simulation that suggests teams should have their worst high-stress players take the first shots, as those are guaranteed to be the least stressful in the shootout. We observe the proportion of shootouts that end after a certain number of kicks, and we conclude that if coaches are able to identify their best high-stress kickers, the first team's best kicker should kick in either round 4 or 5, while the second team's best kicker should kick in either round 3 or 4. We also observe that the shootout structure provides an inherent advantage to the first team to shoot in shorter shootouts and an inherent advantage to the second team in longer shootouts.

JEL classification: C79

Keywords: Soccer, Penalty Kick Shootouts, Optimality, Game Theory

Table of Contents

I. Introduction

Penalty kick shootouts in soccer are perhaps the most exciting tiebreakers in sports, pitting a series of field players against the opposing goalkeeper in a competition involving skill, mental prowess, and nerves. If two teams are tied following 90 minutes of regulation play and after the two 15-minute periods of extra time, each team selects five players to shoot against the goalkeeper from 12 yards away with the team scoring more goals of this sort being declared the victor.[2] This tiebreaker is used for any "knockout" games in international competitions—the FIFA World Cup, the UEFA European Championships, as well as other continental and domestic tournaments— so it is also a high-stakes tiebreaker with financial and professional implications for the winners. In the 2010 World Cup, the winning team received $7m more than the runner-up, and a team that advanced to the quarterfinals received $9m more than a team knocked out in the previous round.[3] Additionally, the World Cup winners in 2002, 1998, 1994, and 1990 all had to win penalty shootouts en route to the championship: during the recent "dynasty" period of the Spanish national team, in which it claimed three consecutive major international competitions from 2008 to 2012, Spain won shootout victories at two separate European Championships.[4] Thus, any advantage in a penalty shootout, however slight, could be very valuable in contributing to that team's international success, development, and reputation.

As the shootout has very quantifiable results—each shot is either a goal or a miss; each shootout has a winning team and a losing team—it is ripe for analysis. While the game theory of penalty kick shot placement—in terms of mixed strategies for both the shooter and the goalkeeper—has been examined and modeled, one area where the current literature is less than satisfactory is in determining the optimal five-player lineups for penalty kick shootouts. Some teams prefer to put their best players first, in an attempt to ensure that they get off to a good start, while some teams put their best players last, for when the stakes are arguably highest. Essentially, the only element that can be agreed upon, identified by both Franks and McGarry (2000) and Macey-Dare (2013) is that teams should select their best five penalty kick takers to kick in the shootout (and their best penalty kick stopper as their goalie)—a finding that is

---

[2] If the teams are tied after the first five players for each team have shot, the shootout continues in a sudden-death fashion, round by round, until a victor is determined.

[3] Gibson, O. (2009, December 3). Prize money increased 61% for 2010 world cup. *The Guardian.* Retrieved from http://www.theguardian.com/football/2009/dec/03/world-cup-2010-prize-money

[4] Slater, S. (2010, May 27). *Penalty shootouts: World cup history*. Retrieved from http://www.fourfourtwo.com/news/penalty-shootouts-world-cup-history

intuitive but leaves the door open for more detailed conclusions. Additionally, it has been suggested [by Billsburry and Nelson (2007) and Macaree (2012)] that penalty kicks are actually not a totally fair tiebreaker, as they may provide advantages to certain nations and may lend an advantage to the first team to shoot.[5] Thus, the opportunity to derive optimal strategies for first-shooting teams and second-shooting teams is intriguing, as teams may be able to further capitalize on an advantage or mitigate a disadvantage.

In this paper, we will analyze the performance of players in past shootouts and develop a comprehensive model for how teams should order their lineups. First, we use a simulation, similar to that of Franks and McGarry (2000), and game theory analysis to provide some general insights and intuitions for the results that follow. Then, we compile an exhaustive database of penalty kick shootouts from international competitions as well as relevant player-specific statistics and analyze this to accurately model player performance in shootouts. From there, we update our simulation with historically accurate numbers to see how teams should order their lineups to maximize their win probability.

However, before proceeding to our work, we briefly discuss relevant literature.

II. Literature Review

Game theory—the study of how competitive players make strategic decisions given a certain situation—has great relevance to sports, and it has been applied to everything from tennis serves to baseball pitch selection to setting basketball lineups. Walker and Wooders (2002) use game theory to test whether professional tennis players are effectively using mixed-strategy equilibrium to maximize their chances of winning.[6] They broke a tennis serve—the initial shot in any point—into simpler, discrete actions consisting of one player serving to the right or left and the other player moving to his right or left. The authors found that the better the player, the more likely they were to effectively use varied serve placement—a mixed strategy—to win games.

Ari Caroline (2012) uses game theory to explain the optimal lineup for a basketball team, taking game theory in sports and looking at it from the coaching perspective. Empirically, "big"

[5] Billsburry and Nelson identify a bias in the manner that individuals from more individualistic nations feel less support and feel more pressure in shootouts and thus perform worse than individuals from more collectivistic nations. Alternatively, more general biases could be demonstrated here: penalty shootouts intuitively favor countries with good goalkeeping and shooting—nations with a relative advantage at passing, dribbling, and defending may be at a disadvantage.

[6] As Walker and Wooders explain, professional sports serve as a good contrast to regular experimental settings, because all players are highly motivated and have a wealth of experience. In professional sports, people play their respective games well, instead of simply playing the games as they would in an experiment.

players on a basketball team (forwards and centers) contribute a higher percentage of production (as measured by points, rebounds, etc.) than guards. Thus, one might expect the optimal strategy for a basketball coach to be putting five “bigs” on the court; however, Caroline suggests that a “small” player could amplify the impact of a “big” through floor stretch—a relative function that has an impact on the probability of a team winning based on what type of lineup the other team fields. In this way, floor stretch models game theory in a basketball lineup setting; both teams will try and set lineups that optimize their likelihood of winning, leading to lineups that we see in a traditional basketball game.

The game theory of penalty kick placement—in a two-player model with the goalkeeper and the shooter—has been examined at length. Baumann and Friehe (2011) use mixed strategy game theory to investigate the best way for kickers and goalies to shoot/jump in a penalty shootout situation. Their model focuses on the kickers’ natural and unnatural side to shoot. This is an interesting economic way to go about investigating penalty kicks but it does not look at the entire series of shots as a whole (as our paper does), focusing instead on each individual goalie-shooter interaction. Similarly, Chiappori and Levitt (2002) studied penalty kicks as a way to test mixed strategy equilibrium. By dividing the goal into a 3-by-3 grid, they measured the likelihood of penalty kick success for shots across these nine quadrants. Ultimately, they found results consistent with the predictions of the game theory model.

These papers consider the strategy for an individual kick in a multi-kick shootout. However, our research focuses on the overall strategy for a shootout. Franks and McGarry (200) explored this topic in their computer simulation of the five-player penalty shootout, and they found that teams in penalty kick shootouts should order players in increasing order of ability, such that they start shootouts with their fifth best kick taker and finish with their best one. Although the researchers develop an intuitive model as their basis for this, their assumptions on player ability, situational stress, and player reactions to situational stress are not supported by empirical evidence. Notably, the researchers assumed that stress increased linearly from the first kick to the fifth kick, a fact that Jordet (2006) has directly refuted. Jordet’s analysis shows that player success declines from the first round of the shootout onward at the aggregate level—players shooting first have the highest conversion rate, which declines as the shootout continues. More interestingly, however, Jordet and others have documented that player performance only substantially declines in situations where he must make a kick for his team to remain alive;

additionally, player performance actually improves when a player can make a kick to win the shootout for his team.

It is worth noting that these findings are not necessarily controlling for player ability or other factors—it is possible that these fluctuations in success rate can be explained by other variables. For example, the players making a kick to win the shootout for their team are often in the fifth slot, where teams may be inclined to put their best player. Alternatively, teams may put their best players first, which partially explains the decline in ability as teams move toward the second through fifth shooters. We hope to confirm that the increases and decreases in penalty kick success rate are not a result of self-selection bias.

The other assumption that Franks and McGarry make that we hope to investigate is how players respond to stress. The model they create assumes that the best players are the most consistent and least sensitive to stress, and the worst players are the most sensitive to stress. Franks and McGarry model this by having player's variance increasing in a linear manner from the best kicker to the fifth best kicker. While Franks and McGarry state that this phenomenon has been observed in other fields, it has not been studied enough to see if it actually takes place during soccer penalty shootouts. Anecdotally, though, the shootout is as much a mental competition as it is skill-based: some teams utilize goalkeepers as penalty kick shooters, and these players have the highest conversion rate for any position.[7]

Given these findings of varying performance in the different "states" that exist in a penalty shootout, a Markov chain-model is of great interest. Bukiet and Harold (1997) used Markov chains in baseball to predict the batting order that maximizes the run output of certain lineups; this same principle can be applied to maximize the win probability of different lineups.[8] For baseball, Bukiet and Harold considered 25 different states, as a combination of the ways runners could be aligned on the bases and the number of outs in the inning. In our case we consider different states based on which circumstances significantly affect the likelihood of the kick's success: in this case, different states are "must-convert," "convert to win," and "general." One type of kick that Jordet doesn't identify is the "semi-must-convert" kick: this is a kick in which if an individual misses, the other team will have a chance to win. We hope to identify performance under this state as well—we hypothesize that performance here will be greater than

---

[7] See summary data.

[8] The fluctuation of win probability in penalty shootouts has been touched upon briefly in popular literature (the Macaree blogpost, most notably), although not in any formal publications.

in the "must-convert" scenario but worse than in the "general" case. By revisiting the Franks and McGarry model and improving upon its precision, we will be able to determine the strategies that maximize win probability within the shootout.

Finally, Billsburry and Nelson (2007) and Macaree (2012) noted that penalty shootouts might not be a completely fair way to decide a tie game, as they could lead to a first or second mover advantage. Ignacio Palacios-Huerta (2012) suggests a way to make sure that either first-mover or second-mover advantage is negated in tournament formats (such as the back and forth shootout). His suggestion, called the Prouhet-Thue-Morse sequence, can prevent both types of advantage and set the most equal possible playing field for participants,. This sequence uses bitwise negation to create a situation that is both ex-post and ex-ante fair. Take the case of the sequence AB. To negate any advantage that may be created from that, you follow it with the inverted sequence BA. Combining those creates ABBA, a pattern that will be fair in the case of any inherent advantage to the first or second player (or team). Continuing to negate the sequence chains in larger and larger groupings will create a series of actions that remains fair throughout the tournament. Although this alternative seems appealing, we hope to investigate the penalty shootout over a greater empirical set to see if there is something inherently unfair about the shootout's current structure.

III. Theoretical Framework

Much of this paper will analyze data of penalty kick shootouts and model player performance with respect to a number of factors: player age, player experience, goal-scoring ability, and the kick situation. Latter sections will apply principals of game theory in order to determine an optimal lineup of players with varying abilities.[9]

First, we mimic the Franks-McGarry analysis with a more rigorous understanding of which kinds of kicks are stressful. We create small simulations to illustrate the role of stress in the shootout, showing that based on the relative stress response of players of different abilities the optimal lineup will shift accordingly.

In the second portion, we model player performance in shootouts using a logit model. Explanatory variables include level of competition (quarterfinals, semifinals, etc.), scenario

[9] Although we could proceed in the opposite order—with the theoretical model coming first—we want to make sure the findings have some basis in reality. We believe, based on previous evidence and intuition, that there's a not insignificant chance that player ability on penalty kicks is entirely random, and thus there may be limited opportunity for strategy optimality from the coach's perspective. With our current setup, we avoid the possibility that we'll devote a significant portion of time to analysis that is ultimately fruitless.

("must-make," "potential winner"—both as described above), a proxy for the player's ability, and a proxy for the goalkeeper's experience. The main proxy we will utilize for ability is a combination of the player's experience and the player's current "form" or goal-scoring ability. The player's experience is measured via his number of appearances (or CAPS) prior to that tournament; the player's current form is measured by how many goals he scored per game in the given tournament. After analyzing the empirical data, though, we see how significant these factors are in the player's success on penalty kicks. Others factors such as player position, captaincy, year, tournament, and minutes played (via a proxy of whether they started the game or entered as a substitute) will be included when modeling the success of each kick, as they were in the Jordet (2006) analysis.

We then take these different factors and regress with respect to "Outcome" (make or miss) to see which has a more significant effect on the player's success rate. We regress using a logit model as the outcome is binary (1=make, 0=miss), and we have no prior belief that the distribution of player ability is normal. Once we identify the proper proxies for ability, we can calculate for the general performance of above-average (one standard deviation above the mean), average, and below-average (one standard deviation below the mean) players.

Most importantly however, we will be able to see how the different levels of players respond in stressful situations. In theory, if players react in identical ways under stress, there is no complicated strategy to organizing a lineup: the manager should simply order his players by way of decreasing ability, such that his best shooters are always guaranteed to kick, thus maximizing the expected scoring output of the lineup. However, if there are varying responses to stress, more complicated lineup strategies could be optimal.

Finally, we hope to be able to provide an empirical estimate for the values of the probabilities of success when under stress and when not, and then we can proceed to identify optimal strategies for each team. In this sense, we will update our hypothetical simulation from the first section with the results from our logit regressions. Since the penalty shootout is a one-off affair—the next opportunity for two teams to engage in a penalty shootout at a major tournament would come two years in the future at the earliest—the goal is to identify a pure strategy for both the first and second teams. If a mixed strategy is optimal for each team, then without guaranteed repeated interaction, the practical utility of this finding, although interesting, is minimal.

Placing one's best players in the highest leverage situations may be the optimal strategy. However, it is likely that the highest leverage situations will appear across a range of different kicks—that is, in one shootout, the fifth kick for a team may be of the highest leverage, while in another, the fourth kick is the most important. The risk aversion of coaches in a certain regard will be interesting to observe empirically; we expect to observe coaches placing their best penalty kick takers early in the shootout (in one of the first three positions), in order to guarantee that the stars get to shoot.[10] In the 2012 European Championship semifinals, Portugal's star player, Cristiano Ronaldo, was slotted to take the fifth penalty kick for Portugal (the tenth kick in the shootout). But as his team was eliminated 4-2 after nine kicks overall, he never got to kick, and Portugal manager Paulo Bento was harshly criticized for this lineup decision.[11] It is unclear whether Bento's strategy was optimal and/or risky, but an investigation of typical placement of a team's best penalty taker, and what strategy is actually optimal, will be intriguing. It is possible that placing the best shooters last will lead to a higher expected output of goals but also a greater variance in the expected output, which coaches may be wary of selecting due to risk aversion.

III.1 Theoretical Simulation

In Table 1 below, we have simulated a 2x2 shootout, using fictitious values for a player's conversion rate in the general and stressful state. The abbreviated simulation allows us to highlight the more general principles at work in the 5x5 shootout. In the chart below, players experience no decline in ability in stressful situations, leading to no variability in expected outcome based on the lineups either manager chooses. Whether the coach of team 1 puts his best player first or second, the expected win probability for his team is constant; the same situation applies for the coach of team 2.

[10] Here we are assuming that coaches "lock in" their lineups prior to the shootout. This is not required by FIFA, but we believe it is a reasonable assumption, as we expect there to be adverse effects from re-arranging a lineup mid-shootout. For the theoretical portions of this paper, we are assuming that these mid-lineup changes do not occur.

[11] O'Mard, M. K. (2012, June 27). *Blame Paulo Bento for Cristiano Ronaldo Not Taking Decisive Penalty Kick at Euro 2012*. Retrieved from http://nesn.com/2012/06/blame-paulo-bento-for-cristiano-ronaldo-not-taking-decisive-penalty-kick/

Table 1: Player 1 Better, No Stress Effects

	Team 2 Best Player 1st			**Team 2 Best Player 2nd**		
	WinProb Team 1	Tie Prob	WinProb Team 2	WinProb Team 1	Tie Prob	WinProb Team 2
Team 1 Best Player 1st	**.2692**	**.4616**	**.2692**	**.2692**	**.4616**	**.2692**
Team 1 Best Player 2nd	**.2692**	**.4616**	**.2692**	**.2692**	**.4616**	**.2692**

Player 1 General Conversion rate = 80%, Player 2 General Conversion rate = 70%
Player 1 Stressful Conversion rate = 80%, Player 2 Stressful Conversion rate = 70%

In Table 2 below, we see a scenario in which players 1 and 2 for both teams perform proportionally worse in stressful situations: here, the success rate of both players declines by 10% under stress. Team 1's dominant strategy is to put its best player first: its win probability is higher when it puts its best player in this slot, regardless of whether team 2 puts its best player first or second. The second team has identical win probabilities for either strategy it chooses (either of the top two quadrants). Here, it will put its best player first to minimize team 1's win probability. Hence, the teams will simply order their lineups starting with their best player and decreasing in ability from there: this quadrant is the bolded strategy below.

Table 2: Player 1 Better, Proportional Stress Effects

	Team 2 Best Player 1st			**Team 2 Best Player 2nd**		
	WinProb Team 1	Tie Prob	WinProb Team 2	WinProb Team 1	Tie Prob	WinProb Team 2
Team 1 Best Player 1st	**.2954**	**.4620**	**.2426**	.3026	.4549	.2426
Team 1 Best Player 2nd	.2926	.4550	.2524	.3006	.4470	.2524

Player 1 General Conversion rate = 90%, Player 2 General Conversion rate = 80%
Player 1 Stressful Conversion rate = 81%, Player 2 Stressful Conversion rate = 72%

This is not the typical assumption, however; both coaches, anecdotally, and scientists assume that better players react relatively better to stress, meaning that their conversion rate drops comparatively less in stressful situations (Franks and McGarry, 2000). For the purpose of this paper (supported by our empirical analysis on page 21), a high-stress kick is any in which a miss by the player will result in a loss for his team or give the other team a chance to win on its subsequent kick. All other kicks are classified as low-stress. The goal for us here is to observe whether optimal lineup orders will change as the relative response to high-stress kicks changes. Here we present different general cases that show this is precisely the case.

When investigating various 2x2 simulations, we have found that teams will structure their lineups in a manner that puts players in a position of comparative advantage. For instance, if the best player on a team is comparatively better under the low-stress state, teams will choose a lineup such that their best player finds himself in relatively more low-stress situations than a random lineup. Below we construct such a situation. Here, team 1 will have its best player shoot first, because that situation is always a low-stress one by construction; team 2 will have its best player shoot second, because that situation is proportionally less likely to be high-stress. This pattern occurs for a variety of different realistic parameter values that we simulated.

Table 3: Player 1 Relatively Better in the General State

	Team 2 Best Player 1st			**Team 2 Best Player 2nd**		
	WinProb Team 1	Tie Prob	WinProb Team 2	WinProb Team 1	Tie Prob	WinProb Team 2
Team 1 Best Player 1st	.4750	.3150	.2100	**.4825**	**.2400**	**.2775**
Team 1 Best Player 2nd	.3700	.2400	.3900	.4000	.1650	.4350

Player 1 General Conversion rate = 90%, Player 2 General Conversion rate = 60%
Player 1 Stressful Conversion rate = 50%, Player 2 Stressful Conversion rate = 50%

Alternatively, when the best player has a comparative advantage in the stressful state, both teams will have their best player shoot first: team 2 chooses to do this because the first kick for team 2 is almost always stressful. Team 1 will put its best player in either the first or second slot depending on how large the comparative advantage is—if it is small, team 1 will have its best player shoot first to attempt to prevent a high-stress situation from arising; if it is large, team 1 will have its best player shoot second, in the slot that is more likely to be high-stress. Table 4a outlines the first situation; Table 4b outlines the second.

Table 4a: Player 1 Relatively Better in the Stressful State: Small Advantage

	Team 2 Best Player 1st			Team 2 Best Player 2nd		
	WinProb Team 1	Tie Prob	WinProb Team 2	WinProb Team 1	Tie Prob	WinProb Team 2
Team 1 Best Player 1st	**.3265**	**.3825**	**.2910**	.3550	.3675	.2775
Team 1 Best Player 2nd	.3220	.3708	.3072	.3496	.3468	.3036

Player 1 General Conversion rate = 90%, Player 2 General Conversion rate = 60%
Player 1 Stressful Conversion rate = 80%, Player 2 Stressful Conversion rate = 50%

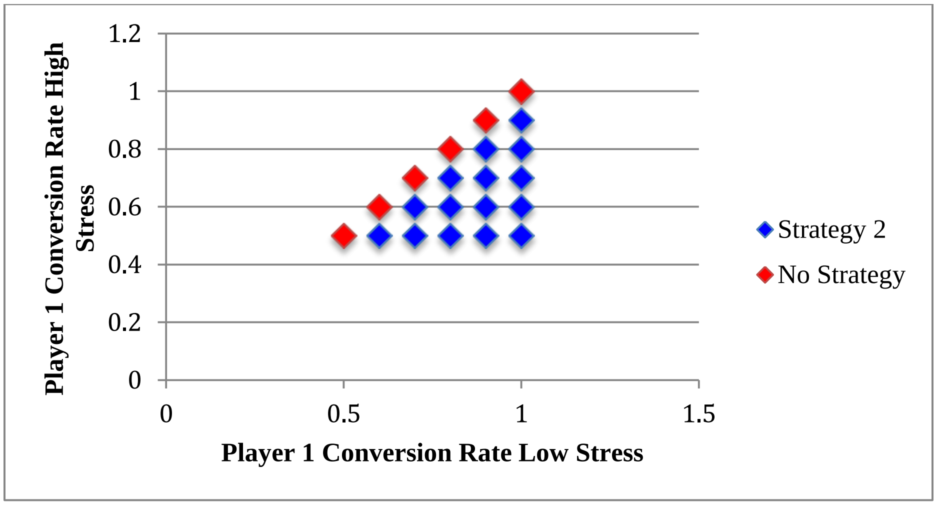
Table 4b: Player 1 Relatively Better in the Stressful State: Large Advantage

	Team 2 Best Player 1st			Team 2 Best Player 2nd		
	WinProb Team 1	Tie Prob	WinProb Team 2	WinProb Team 1	Tie Prob	WinProb Team 2
Team 1 Best Player 1st	.3355	.3325	.3320	.3900	.3275	.2825
Team 1 Best Player 2nd	**.3690**	**.3432**	**.2878**	.3924	.3352	.3724

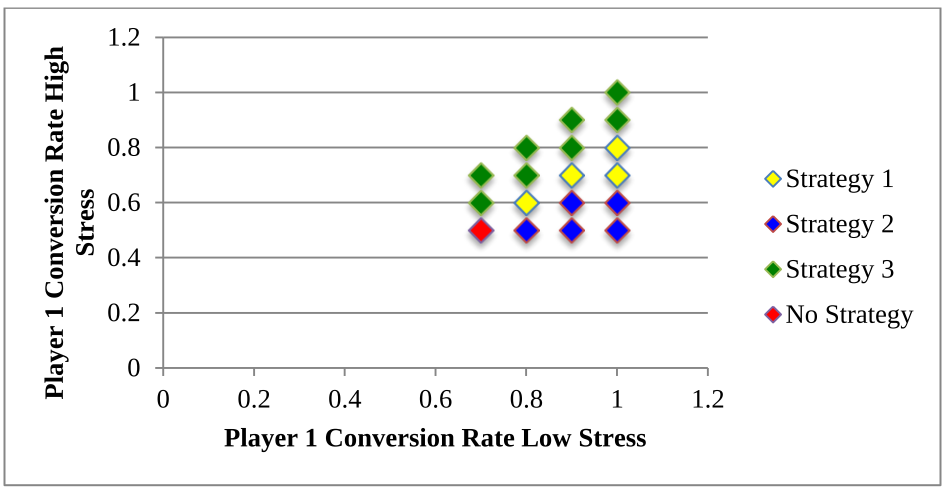
Player 1 General Conversion rate = 90%, Player 2 General Conversion rate = 70%
Player 1 Stressful Conversion rate = 80%, Player 2 Stressful Conversion rate = 50%

The three graphs below illustrate a more exhaustive investigation of different theoretical parameters for players in the 2x2 model. In each chart, player 2 is held constant at different values while we change the high-stress and low-stress conversion rates of player 1. The graphs contain all incremental values until player 1 is only just as good as player 2. Strategy 1 corresponds to both teams putting their best player first; Strategy 2 has the best player for team 1 first and the best player for team 2 second; Strategy 3 has the best player for team 1 second and the best for team 2 first; Strategy 4 has the best player for both teams going second.

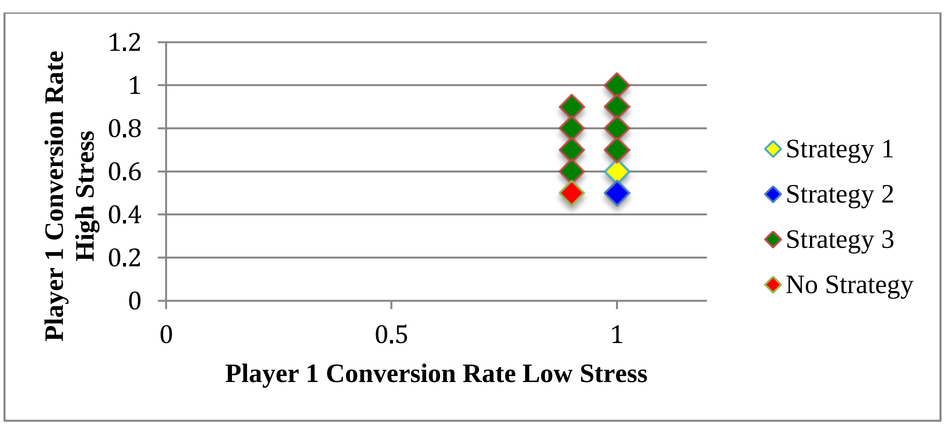
Sensitivity Test Graph 1a - Player 2 Conversion Rates (.5 in low stress, .5 in high stress)



Sensitivity Test Graph 1b - Player 2 Conversion Rates (.7, .5)



Sensitivity Test Graph 1c - Player 2 Conversion Rates (.9, .5)



In graph 1a, player 2 is held at a constant 50% conversion rate in both the high- and low-stress states. When player 1 has no comparative advantage in the high- or low-stress state (along the 45-degree diagonal) order does not matter. If there is a difference in conversion rate between the high- and low-stress state for player 1, Strategy 2 is optimal, since player 1 has a comparative advantage in the low-stress state. Thus, team 1 puts player 1 first, and team 2 puts him second.

Graphs 1b and 1c hold constant player 2's conversion rates in the low- and high-stress states at (70%, 50%) and (90%, 50%), respectively. Whenever player 1 and player 2 are of equal ability, the order does not matter, as each player is interchangeable. When the players have roughly identical proportional responses to stress, Strategy 1 is optimal—that is, players are ordered in decreasing order of ability for both teams. When Player 1 has a comparative advantage in either the high- or low-stress states, more strategies come into play: strategy 2 is utilized when the comparative advantage is in the low-stress state, and strategy 3 is employed when the advantage is in the high-stress state. In Chart 1b, Strategy 2 or Strategy 3 is employed at different values as the comparative advantage shifts.

In Graph 1c, player 2 has a high conversion rate in the general state. This leads to Strategy 3 being the most useful, as team 1 will want to put player 1 in his spot of comparative advantage (high stress, which usually can only happen on the second shot for them). Similarly, team 2 will want to put player 1 in his high-stress shot, which is often the first shot for that team.

These graphs more exhaustively illustrate the points from the tables above: coaches will want to put their best player in his place of comparative advantage whether it be the high- or low-stress shot. Similar findings of varying lineups based on the comparative advantages of players have been confirmed in 3x3 simulations as well. These findings are listed in the appendix. With this thesis, we hope to be able to provide an empirical estimate for the values of these probabilities, and then we can proceed to identify optimal strategies for each team.

IV. Research Plan

We have collected historical data on shootout performance for teams and individuals. For each individual, we identify two proxies for ability: 1) the individual's experience prior to the tournament, measured via his number of appearances (CAPS) for the senior national team, and 2) the individual's goal-scoring ability, measured via his number of goals scored during the tournament in question (to measure his current "form"). We use this measure of ability to rank players for each team and observe how players of different rankings perform, and how they perform under situations of different stress.

Currently, the Franks and McGarry model is the most complete simulation of a penalty kick shootout, but their understandings of player responses to stress were inadequate. Additionally, their assumptions of distribution of penalty-kick taking talent were not based in reality. Through our empirical section, we develop more accurate understandings of both of these. We use these empirical findings to mimic actual shootout performance in our game theoretical simulations. We identify the average success rates for players of different abilities in both high- and low-stress situations and we use these values as the parameters in our model. From here, we determine the optimal lineup order for both the first and second teams .

Finally, we investigate how the first-team advantage in penalty shootouts holds up over a high volume of shootouts in our analysis. We calculate the likelihood of this higher win probability for the first team to shoot given an equal initial win probability for each team and see if it is significant, given an expectation of the shootout as a fair tiebreaker. It is our hypothesis that the team subject to fewer high-pressure shots will have an advantage overall. Franks and McGarry predicted an advantage for the second team; however, this assumption stemmed from their use of only 18 shootouts upon which they extrapolated initial trends. Studies that include a greater set of shootouts typically show the first team to shoot winning 60% of the shootouts; however, these data sets are also incomplete.[12]

V. Data

Data from individuals were collected from all international penalty shootouts in elimination games. In total, this data set is comprised of 98 shootouts across seven international competitions: FIFA World Cup (1982-2010), Confederation Cup (1995-2013), European Championships (1972-2012), Copa America (1993-2013), CONCACAF Gold Cup (1991-2013), Africa Cup of Nations (1988-2013), Asian Cup (1996-2011). The data set includes 1014 penalty kicks. Seven additional shootouts were included solely for the breakdown of long and short shootout win percentage for first and second teams in Table 14—in these shootouts, it was clear which team shot first, but not the individual breakdown on the shots. Data have been compiled

[12] Kuper, S., & Syzmanski, S. (2009). *Soccernomics: Why England Loses, Why Spain, Germany, and Brazil Win, and Why the US, Japan, Australia, Turkey-and Even Iraq-are Destined to Become the Kings of the World's Most Popular Sport*.

from online databases such as worldfootball.net, 11v11.com, national-football-teams.com, and corroborated with the Wikipedia entries for these games.

To measure ability, proxies were taken via CAPS prior to the tournament and goals scored during the tournament. Both of these tallies are pulled from the same databases cited above, as well as tournament-specific archives from rsssf.com. Player age was also obtained to allow for consideration of natural increases and decreases in ability over time in our analysis. We expect more experienced players and more in-form players (who have scored more in the current tournament) to have a higher success rate; additionally, we expect players' ability to peak over the middle of the careers and decline at the start and end of their careers. To control for variability in player responsibilities, which would contribute to biases in scoring production, we also gathered positional data—identifying players based on how offensive / defensive a position they maintained in their careers for their national teams.

The other primary data usage is the identification of the scenario for each kick. This is again drawn manually from the penalty kick order (from Wikipedia or worldfootball.net). The scenario classifications are determined manually as follows: if a successful kick will win the shootout, that kick is a "clinching" kick; if a missed kick will lose the shootout, that is a "must-make"; if a missed kick will lead to a "clinching" kick for the other team, that first scenario is a "semi-must-make." All other kicks are classified as "general." The clinching and must-make kicks have been identified as significant differentiators in previous literature (Keh 2010, via Jordet 2006). The concept of the semi-must-make, however, is a novel one.

VI. Results

We observe the following conversion rate breakdowns by shot number: 79.6% (156/196) on first shot, 79.1% (155/196) on second shot; 70.4% on third shot (138/196); 68.4% on fourth shot (128/187); 79.3% on fifth shot (96/121). These percentages are relatively similar to Jordet's (2006) findings over a smaller data set. This distribution does fall somewhat in line with the placement of the best shooter in shootouts: when ranking the players based on their career statistics (measured in an ex-post manner) or their experience at the time, we see that over 70%

of coaches elect to put their best player in one of the first three slots, guaranteeing that they will kick in the shootout.[13]

Table 5a-Placement of Best Shooter In Shootout (Career Goals)

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Total
Occurrences	74	41	39	26	21	201
Percent of Overall	36.8%	20.4%	19.4%	12.9%	10.4%	100%
Cumulative %	36.8%	57.2%	76.6%	89.5%	99.9%	100%

Table 5b-Placement of Best Shooter In Shootout (Career CAPS)

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Total
Occurrences	62	35	37	25	23	182
Percent of Overall	34.1%	19.2%	20.3%	13.7%	12.6%	100%
Cumulative %	34.1%	53.3%	73.6%	87.3%	99.9%	100%

Table 5**c**-Placement of Best Shooter In Shootout (Prior CAPS)

	Shot 1	Shot 2	Shot 3	Shot 4	Shot 5	Total
Occurrences	60	38	40	28	24	190
Percent of Overall	31.6%	20.0%	21.1%	14.7%	12.6%	100%
Cumulative %	31.6%	51.6%	72.7%	87.4%	100%	100%

At the aggregate level, we have observed that players convert 76.1% of general kicks (504 of 662 attempts) and 78.2% of clinching kicks (61/78 attempts); players convert slightly less frequently on must-make and semi-must-make kicks—69.2% (81/117 attempts) and 71.3% (112/157 attempts), respectively. This illustrates what has been intuitively assumed both in the above tables and in our research in general; the "high-stress" scenarios (must-make and semi-must make) have a lower conversion rate than the "low-stress" scenarios (general and clinching). To further delineate this, we regress the outcome of the shot (either make or miss) on whether it was taken in a stressful or non-stressful scenario.[14]

[13] Again, if the best shooter for a team does not get to kick because the shootout ends prior his scheduled attempt, as in the famous Christiano Ronaldo situation at Euro 2012, then the second-best shooter will be ranked as the best, the third-best as the second-best, and so forth.

[14] This is not the full regression: Table 6 is simply included for illustrative purposes.

As stated above, we classify these kicks as either high-pressure (must-make and semi-must-make) or low-pressure (general and clinching). Assuming a one-sided t-test, we observed a significant decline in conversion rate from low-pressure kicks to high-pressure kicks at the 5% significance level (t=-1.925, p=.027).

Table 6 - Significance of High-Stress

Variable	Coefficient (Std. Deviation)	P-Value
High-Stress Scenario Dummy	-.3038 (-.1518)	.055
Constant	1.172 (.0865)	0.00

The data produce the following set of logit regressions, shown in Table 7 with the actual coefficients from the regression. The marginal effects from the third regression appear in Table 7d. The high-stress variable is statistically significant at either the 5% or 1% level across all regressions. Additionally, when considering our common explanatory variables (age, age squared, goalkeeper experience) across all regressions, we observe significant, useful patterns: shooters are less likely to convert when the opposing goalkeeper has more experience, and more likely to convert when they (the shooters) are more experienced. Interestingly, form, as measured by goals scored per game in the current tournament (Tourney Goals Per Game), does not play a significant role in success.

Table 7 - Logit Regressions on Outcome[15]

	(1)	(2)	(3)
	Outcome	Outcome	Outcome
High-Stress	-0.429**	-0.436***	-0.525***
	(0.168)	(0.168)	(0.171)
World Cup			-0.441**
			(0.191)
Age	-0.662**	-0.663**	-0.673**
	(0.272)	(0.272)	(0.283)
Age Squared	0.0114**	0.0113**	0.0116**
	(0.00498)	(0.00498)	(0.00521)
Year	-0.0191**	-0.0196**	-0.0267***
	(0.00945)	(0.00945)	(0.0101)
Tourney Goals Per Game	0.214		0.224
	(0.361)		(0.405)
Scored In Tournament		-0.0523	
		(0.158)	
Experience	0.00239	0.00296	0.00316
	(0.00397)	(0.00397)	(0.00409)
Goalkeeper Experience	-0.00392**	-0.00388**	-0.00364*
	(0.00195)	(0.00195)	(0.00199)
Defender			0.0332
			(0.239)
Midfielder			-0.267
			(0.225)
Defensive Midfielder			0.196
			(0.322)
Attacking Midfielder			0.460
			(0.280)
Constant	49.00**	50.18**	64.56***
	(19.48)	(19.51)	(20.74)
Observations	963	963	949
Pseudo-R^2	0.0228	0.0226	0.0371

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Regression 1 is the only regression where we used tournament goals per game to measure a player's form at the time of their shot. This proved to be statistically insignificant. In the second regression we replaced the Tourney Goals Per Game variable with a binary for whether

[15] 49 observations were dropped due to lack of experience data; two others were dropped due to lack of age data. In regression 3, Stata removed the 14 observations of goalkeepers, because they had converted all 14 of these kicks and thus the goalkeeper dummy variable perfectly predicted success.

the player had scored in the given tournament (since almost 700 players had not scored in the relevant tournament for this metric). However, no effect was detected here, either: recent goal-scoring does not appear to be a strong predictor of penalty kick success.

The other three explanatory variables yield some interesting findings. We expected to see players peak in the middle of their careers and perform worse at the beginning and end. The data shows the opposite: players are actually more successful when they are younger. One plausible explanation for this observation is that younger players are unknowns to the goalkeepers: they have no established pattern of shooting behavior on penalty kicks, and thus they have an informational advantage on the goalkeeper.[16] This advantage declines as players become more well-known; by the end of their careers, these shooters have adapted and once again retain their advantage (in what is often termed "veteran guile" or "craftiness").

Another possible explanation is the positive correlation between age and experience: the two variables have a correlation coefficient of .5430, meaning that older players have more career CAPS. This could imply that the Age and Age Squared variables are swallowing some of the effect of the increase in experience, leaving experience to be statistically insignificant. In table 7b below, we ran regressions using each of these proxies—1) Experience and 2) Age and Age Squared—in isolation. The findings do not substantially affect our assumptions made above: although age is no longer significant, it is on the cusp of being significant (p=.115) at the 10% level. Experience alone does not appear to be a good proxy for ability at the aggregate level; however, in Table 7a, we see that when controlling for experience younger players perform significantly better, which suggests that these players are simply more precocious and more talented. We visit the interaction between experience level and stress response later on.

In the third regression, we accounted for the differing positions players occupied on the field. However, no significant findings show up with regard to position. More interestingly, we included here a proxy for playing in the World Cup, which we expected would add another layer of stress to the shootout. Indeed, this environment did make shooters significantly less successful; players in the World Cup had an 8.2 percentage point decrease in their conversion rate relative to any other tournament. In another regression (not shown here) we also included

[16] In multiple high-profile shootouts, goalkeepers have been observed studying notes on shooting habits of opposing penalty-kick takers. In the 2006 World Cup, Portugal's GK Ricardo stopped three penalties against England after studying player habits; in the same tournament, Germany's Jens Lehmann also studied his notes en route to stopping two of Argentina's penalties.

proxies for 1) the round in which the shootout occurred (quarterfinal, semifinal, final); 2) if the player was a substitute (and presumably fresher); and 3) if he was a captain (and presumably more confident). These factors had no significant impact on outcome.

Table 7b - Logit Regressions on Outcome

	(1)	(2)
	Outcome	Outcome
High-Stress	-0.400**	-0.422**
	(0.163)	(0.167)
Year	-0.0185**	-0.0204**
	(0.00923)	(0.00972)
Tourney Goals Per Game	0.274	0.344
	(0.350)	(0.360)
Age	-0.382	
	(0.242)	
Age Squared	0.00641	
	(0.00444)	
Goalkeeper Experience	-0.00364*	-0.00379*
	(0.00190)	(0.00196)
World Cup	-0.353*	-0.407**
	(0.185)	(0.189)
Experience		-0.000166
		(0.00332)
Constant	44.07**	42.28**
	(18.83)	(19.43)
Observations	1,008	965
Pseudo-R^2	0.0202	0.0183

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7c below shows the distribution of age throughout the data set. The mean age is 26.54 with a standard deviation of 3.65. Because only full years were used when measuring age, anyone over 30 falls outside of one standard deviation, as does anyone under 23. Because of the nonlinearity of the regression, at age 59, player ability is predicted to increase. However, it is reasonable to assume that no 59-year-olds will be taking the field for these games as the oldest player in the data set is 40.

Table 7c - Age Distribution

Age	Frequency	Percent	Cumulative %
17	1	0.10	0.10
18	6	0.60	0.69
19	10	0.99	1.69
20	19	1.88	3.57
21	44	4.37	7.94
22	55	5.46	13.39
23	81	8.04	21.43
24	89	8.83	30.26
25	108	10.71	40.97
26	102	10.12	51.09
27	105	10.42	61.51
28	97	9.62	71.13
29	74	7.34	78.47
30	66	6.55	85.02
31	53	5.26	90.28
32	39	3.87	94.15
33	25	2.48	96.63
34	19	1.88	98.51
35	10	0.99	99.50
36	2	0.20	99.70
38	2	0.20	99.90
40	1	0.10	100.00

Table 7d – Conditional Marginal Effects

	Dy/Dx	Std. Err	Z Score	P>Z	95% Con	95% Con
Highstress	-.0811756	.0299762	-2.71	.007	-.1399278	-.0224233
World Cup	-.0824189	.034219	-2.41	.016	-.1494868	-.0153509
Year	-.0047709	.00181	-2.64	.008	-.0083158	-.0012233
Age	-.0077187	-.0047086	-1.64	.101	-.0169473	.0015099
Age Squared	3.34 e-^-6	4.52 e^-6	.74	.461	-5.53 e^-6	.000015099
Tourney GPG	.0259814	.0746856	.35	.728	-.1203997	.1723624
Experience	.0008256	.0007417	1.11	.266	-.0006282	.0022793
GK Experience	-.0007088	.0003547	-2.00	.046	-.1494868	-.0000135
Captain	-.0427225	.0638479	-.67	.603	0.167862	.0824171
Substitute	.0298884	.0638479	.84	.3999	-.0395136	.0992903
Defender	-.0091558	.044453	-.21	.837	-.096282	.0779705
Midfielder	-.065927	.0403132	-1.64	.102	-.1449395	.0130855
D. Mid	.0166626	.058253	.29	.775	-.0975112	.1308364
A. Mid	.0601204	.0496436	1.21	.226	-.0371793	.15742

Table 7d takes the logit regression coefficients from regression 3 of Table 7a and turns them into actual values that can be interpreted for discussion in regards to the likelihood of a penalty shot being converted. High-stress situations and being in the World Cup have the highest negative impact on the likelihood on kick conversion, with and 8.1 and 8.2 percentage point drop coming from those two variables. Tournament Year has a statistically significant result but a much smaller impact, showing only a .4 percentage point drop in conversion rate per year.

We next look to see if players of different abilities have different responses to stress. It is expected that more experienced players would perform better under stress, as they are more accustomed to these situations. Players were sorted based on their relative experience: players one standard deviation above the average (with 50 or more CAPS) were classified as experienced players, and players one standard deviation below the average (with 5 CAPS or fewer at the time of the shootout) were classified as inexperienced. Players with 6-49 CAPS were classified as average. From there, we observed the stress effects for these given sets of players, which are shown in Table 8 below:

Table 8 - Effects Breakdown by Experience Level

	Inexperienced Outcome	Average Outcome	Experienced Outcome
High-Stress	-0.699*	-0.340	-0.439
	(0.381)	(0.207)	(0.474)
Age	-0.591	-0.570	-1.617
	(0.669)	(0.375)	(1.338)
Age Squared	0.00990	0.00968	0.0277
	(0.0132)	(0.00685)	(0.0225)
Goals Scored Per Game	-0.266	0.172	0.670
	(1.336)	(0.426)	(0.896)
Goalkeeper Experience	-0.00463	-0.00490**	-0.00335
	(0.00539)	(0.00238)	(0.00458)
World Cup	0.341	-0.430**	-0.250
	(0.554)	(0.217)	(0.434)
Constant	9.971	9.864*	24.85
	(8.445)	(5.095)	(19.92)
Observations	169	634	160
Pseudo-R^2	0.0365	0.0249	.0237

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In these charts we see that high-stress kicks only have a statistically significant effect on inexperienced players, while experienced and average players both show no significant adverse effects due to stress. We ran linear regressions on outcome and high-stress to see the baseline rates for success in these three experience groups; then, using a conditional marginal effects analysis to determine the decline in success under stress for each group, we were able to determine the conversion rates in both high-stress and low-stress situations for these different levels of experience. These findings are below in Table 9a. Table 9b only considers different conversion rates for the different stress states for the inexperienced players, as only this group experienced significant adverse effects.

Table 9a - Parameter Values

	Low-Stress Conversion %	High-Stress Conversion %
Experienced (1)	76.52	67.86
Average (2)	76.97	71.08
Inexperienced (3)*	77.68	67.24

*Decline in ability is significant at 10% level

Table 9b – Adj. Parameter Values

	Low-Stress Conversion %	High-Stress Conversion %
Experienced (1)	76.52	76.52
Average (2)	76.97	76.97
Inexperienced (3)*	77.68	67.24

*Decline in ability is significant at 10% level

Table 9c –Number of Shooters By Group (%)

	Low-Stress Conversion %	High-Stress Conversion %
Experienced (1)	132 (82.5%)	28 (17.5%)
Average (2)	469 (73.85%)	166 (26.15%)
Inexperienced (3)	112 (65.88%)	58 (34.12%)

Table 9d – Number of Shooters By Group in First 5 Shots (%)

	Low-Stress Conversion %	High-Stress Conversion %
Experienced (1)	128 (87.07%)	19 (12.93%)
Average (2)	461 (80.03%)	115 (19.97%)
Inexperienced (3)	109 (77.86%)	31 (22.14%)

In table 9c, we see that inexperienced players show up proportionally more in high-stress situations than do their average and experienced counterparts. However, one potential reason for this is that longer shootouts are of proportionally higher stress—in sudden-death, all of team 1's kicks are high-stress, and if team 1 converts its kick, team 2's kick will also be of high-stress. Rounds 6-11 in the shootout will involve the less capable shooters (who aren't one of the five best, presumably), which would likely also be where the inexperienced players show up. In table 9d we control for this potential bias and only look at the breakdown within the first five shooters for each team, the initial part of the lineup in which the coach has the greatest opportunity for strategy. Still, the lineups are ordered suboptimally: inexperienced players show up significantly more frequently in high-stress situations than experienced players.

VII. Implications of Empirical Work For Optimal Shootout Lineups

In designing an optimal lineup, we expect "inexperienced" players to appear disproportionately in low-stress situations, since they respond the worst to high-stress situations. In tables 10a and 10b, that is exactly what we see in a 3x3 simulation with both teams boasting an experienced player, an average player, and an inexperienced player: both teams choose to put their inexperienced player in the first slot in equilibrium, ensuring a low-stress kick for these players.[17] Similarly, both teams put their average player in the slot that is most likely to be high-stress (the third kick for team 1; the second kick for team 2), as this player has the best stress response. The equilibrium strategy is in bold below.

Table 10a – 3x3 Empirical

	123[18]	132	213	312	231	321
123	.353 / .343	.353 / .337	.362 / .339	.346 / .339	.362 / .337	.345 / .344
132	.359 / .327	.358 / .322	.369 / .324	.352 / .324	.368 / .322	.352 / .327
213	.352 / .344	.352 / .338	.362 / .340	.346 / .341	.362 / .338	.345 / .344
312	.359 / .326	.358 / .321	.369 / .323	.352 / .323	.368 / .321	**.352 / .326**
231	.355 / .338	.355 / .332	.365 / .334	.348 / .334	.364 / .332	.348 / .334
321	.355 / .336	.355 / .330	.365 / .332	.349 / .332	.364 / .330	.348 / .336

Values: Team 1 WP / Team 2 WP ;; Team 1 Vertical; Team 2 Horizontal [19]

[17] Table 10a considers the parameter values listed in Table 9a; Table 10b considers the parameters in Table 9b. The high-level findings are consistent regardless of which one is considered.

[18] A "123" order signifies a lineup in which the best player shoots first, followed by the second-best player, followed by the third-best. Another illustration:, a "312" lineup means the third-best shoots first, followed by the best, followed by the second-best.

Table 10b – 3x3 Empirical – No Stress Effects for Experienced and Average Players

	123	132	213	312	231	321
123	.332 / .348	.333 / .334	.332 / .348	.293 / .348	.333 / .334	.293 / .348
132	.347 / .310	.342 / .300	.347 / .310	.304 / .310	.342 / .300	.304 / .310
213	.332 / .348	.333 / .334	.332 / .348	.293 / .348	.333 / .334	.293 / .348
312	.348 / .304	.343 / .294	.348 / .304	**.304 / .304**	.343 / .294	**.304 / .304**
231	.347 / .309	.342 / .299	.347 / .310	.304 / .310	.342 / .299	.304 / .310
321	.348 / .304	.343 / .294	.348 / .304	**.304 / .304**	.343 / .294	**.304 / .304**

Values: Team 1 WP / Team 2 WP ;; Team 1 Vertical; Team 2 Horizontal

Although the 3x3 matrix simulation cannot give us specific insights into the 5x5 strategy, we can still make some informed generalizations. The players who experience a negative stress response should go in low-pressure situations (shots one and two), whereas players with the best stress response should go in a position that will almost certainly get to kick but that is also expected to be high-stress. Note here that although we may not see which factors directly play a role in optimal stress response, the individual coach for each team can identify which of his players responds the best under stress (through a nuanced understanding of each player's strengths and weaknesses). As long as the coach can identify these stress responses, he can make an optimal lineup given his pre-shootout information.

Although we do not formally model or empirically test the risk aversion of the coach, our results can be extended to consider these risk attitudes. High-stress kicks appear at different times during the shootout—they can appear anywhere from the third kick for Team 1 onward, depending on whether early shooters convert or miss. Only the first three shooters for each team are guaranteed to kick: regardless of whether a team is shooting first or second, it can win/lose the shootout prior to the fourth round of kicks. The historical patterns for the presence of high-stress kicks and the length of shootouts are listed below in Tables 11 and 12.

In Table 11, assuming that the best player has the best response to stress, we see that the second team should likely put its best player one slot ahead of where the first team should.[20] In a 5x5 shootout, this would likely mean the best shooter goes fourth for team 1 and third for team 2. If the coaches are more willing to take the chance that their best player does not get to kick, a secondary equilibrium could arise with the best shooter shooting fifth for team 1 and fourth for

[19] These values do not add up to 1 because there is a probability that a given shootout is tied after three kicks: for presentation purposes, we did not include the probability of a tie.
[20] When "best player" is used subsequently, it is assumed that this player responds best to stress.

team 2, as these specific shots (numbers 8 and 9 in the aggregated shootout order) are the most likely to be high-stress.

Table 11 – Distribution of High-Stress Shots

Shot #	5	6	7	8	9	10	11	12	13	14
# High	3	16	31	47	49	27	23	12	13	12
# Shots	98	98	98	89	75	46	23	23	13	13
% HS	3.06%	16.33%	31.63%	**52.81%**	**65.33%**	58.70%	100%	52.17%	100%	92.31%

Table 12 - Distribution of Shootout Length

Ending in	7 shots	8 shots	9 shots	10 shots	11+ shots	Total
#	9	14	29	23	23	98

This riskier strategy makes more sense for team 2, which historically has about a 9.2% chance that its fourth player will not get to kick (since 89 of 98 observed shootouts have continued through that kick—only 9 have ended before that). Team 1 would be slightly more risk-tolerant to put its best player fifth, as there is a 23.47% chance this fifth player would not kick. Although technically there is a chance a shootout could end after six shots (a 3-0 win for either team), that never happened in the observed data set. Therefore, Team 1 should not play its best player prior to the fourth position, as both its third and fourth kickers are almost certainly guaranteed to kick and the fourth kick is more likely to be of the high-pressure variety by roughly a factor of 10 (31.63% of fourth kicks for Team 1 are high-pressure, while just 3.06% of third kicks for Team 1 are).

In an analysis of the empirical approach of coaches, there does not appear to be a significant trend in terms of strategic placement of players. 147 above-average players appeared in the first five shots for either team in our data set (i.e. in the lineup for a regular-length shootout): of these 147, 113 (76.9%) were placed in one of the first three positions, guaranteeing their opportunity to kick. Similarly, for the below-average players, 60% of these players in the first five were placed in one of the first three (guaranteed) shooting positions. Our analysis suggests that a more nuanced strategy may be necessary: above-average players, or those with optimal stress responses, should be placed in positions that are disproportionately likely to be of high-stress. However, since it is impossible to measure each individual player's stress response, we cannot say for certain whether these lineups have been determined sub-optimally, although a very preliminary analysis seems to suggest this.

When interpreting the regressions, it is important to note that the stress variables have a substantially larger impact on a shooter's conversion rate than any proxies we have used for ability. Referring back to Table 7d for example, one can see that being in a high-stress situation or the World Cup leads to a drop of 8.7 and 8.2 percentage points, respectively, in the shooter's ability to convert. In comparison, each additional CAP prior to the tournament only has a .08 percentage point increase in the shooter's conversion rate. This implies that it is more important for a coach to understand which of his players respond better or worse to stress and act accordingly. Because these are one-off events with high variance, ability does not have as large an impact on success as the player's response to the stress of the situation. Were this analysis to be continued, the most crucial element would be determining a better proxy for penalty kick-taking ability. Perhaps the success rate of players' penalty kicks taken in club competition could be ascertained; however, such a manual undertaking was unfeasible given the timeframe of this project and the lack of a de facto relevant database. Hopefully future analyses can build upon the dataset created here and begin compiling more player-specific variables to predict success in shootouts.

The slight negative impact of the year of the tournament on the likelihood of conversion is interesting. One explanation for this goes hand-in-hand with the statistically significant impact of goalkeeper CAPS on conversion rate. It is highly possible that goalkeepers have improved more over the past 40 years than kickers have, as techniques have changed and the emphasis on the goalkeeper as a position has become larger. Throughout soccer's history, scoring has traditionally been emphasized, so perhaps shooters reached their ceiling of shot conversion before goalkeepers reached their ceiling of shot-blocking. Improved technology in goalkeeper equipment also may have a similar effect. Additionally, goalkeepers may now realize that they are able to leave the goal line slightly earlier than is legal in shootouts and cut down the angle on shooters.[21]

Another explanation could be that as time has progressed and the stakes of the shootout have increased, the stress of the situation has had a larger impact on each player. As we noted in the introduction, the difference between the winner and runner-up in a World Cup tournament was millions of dollars in 2010, with total prize money for participants topping $420 million, a

[21] On a penalty kick, it is illegal for the goalkeeper to move forward off the goal line before the shooter kicks the ball. Still, such a violation is rarely called unless especially egregious, allowing goalkeepers a slight chance to boost their odds of blocking the shot.

61% increase over 2006. In 2014, total prize money will be $575 million, a 37% increase from the previous edition. If this trend were to hold for previous tournaments, it could explain the small decrease in conversion rate as with increased money comes increased stress.[22]

Finally, one point of note is the advantage that the first team holds in short shootouts versus the second team's advantage in longer shootouts. By construction, the first team should have the advantage in "incomplete shootouts" (shootouts in which not all ten players kick). This is because, assuming all players convert the first rounds of kicks, the first high-pressure kick will be the fourth kick for team 2; if team 2 misses this kick, then team 1 can complete an entire shootout without facing a high-pressure kick. If we assume an extreme scenario in which every player makes low-stress kicks and every player misses high-stress kicks, then team 1 will win every shootout. In Table 13, we see Team 2 faces early high-stress kicks roughly 25% more frequently than Team 1. It's important to note that here we are assuming the teams are of equal ability entering the shootouts, which may not be the case and could play into the varying advantages shown in Table 13. Thus, while there are likely multiple factors contributing to the differences in win percentage here, these trends are still interesting to document.

Table 13 – Empirical Distribution of High-Stress Shots By Team

	First Team	Second Team
# High-Stress Kicks	142	133
# Kicks	526	488
% Of All Kicks	27.00%	27.25%
High-Stress % Of Early Kicks (Rounds 3-5)	30.63% (83 / 271)	38.63% (90 / 233)
High-Stress % Of Later Kicks (Rounds 6+)	100% (59 / 59)	72.88% (43 / 59)
Win % in Short Shootouts	58.90% (43 / 73)	41.10% (30 / 73)
Win % in Long Shootouts	46.875% (15 / 32)	53.125% (17 / 32)

Conversely, in long shootouts—shootouts in which teams are tied after 4 rounds of kicks and are thus entering "sudden death rounds" in which they can directly impact the outcome—the second team has the advantage. Every kick for Team 1 is now high-pressure, because each kick is of the "semi-must-make" variety. Team 2 only enters a high-pressure situation when Team 1 makes its kick; otherwise Team 2 will be in a low-pressure clinching scenario—as shown in

[22] Another explanation may be the tournaments in our dataset whose data go back the furthest. In the sample, the Euro Cup has data going back the furthest. As can be seen in Table VIII in the appendix, the Euro Cup has one of the highest overall conversion rates for a given tournament. Because this tournament is the only data for those early years, this disparity would lead to a higher overall conversion rate then. As other tournaments are added to the set in later years, their lower baseline conversion rates drag down the average and lead to the effect that is picked up in the year variable.

Table 13, only 73% of Team 2's kicks were actually high-stress. This flipping of the first-team and second-team advantage is somewhat reflected in the data: in short shootouts (ones that were not tied after four kicks each), the second team won just 30 of 73 shootouts (41.10%). In long shootouts, the second team won 17 of 32 shootouts (53.125%). Although the p-value for a one-sided test here is insignificant (p-value=.14), we expect that over a greater period a significant difference would develop between win likelihoods over short and long shootouts. Even though at a macro level neither team holds an advantage prior to the shootout, there are certain advantages in the sub-game of each shootout. This could be systematically avoided if FIFA were to switch to an organization of the shootout along the lines of the Prouhet-Thue-Morse sequence as suggested by Palacios-Huerta.

VIII. Appendix

Table I – Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Player	0				
Team	0				
Year	1014	1999.803	8.404069	1976	2013
Competition	0				
Level	1014	2.344181	1.048141	0	4
TourneyGames	1014	2.432939	1.010771	1	5
Outcome	1014	.7475345	.4346412	0	1
Scenario	1014	-.3086785	.7736936	-2	1
GoalsTourney	1014	.7100592	1.170026	0	7
Shot	1014	3.369822	2.054133	1	12
TotalTourn~s	1014	5.327416	1.079879	2	8
TourneyGPG	1014	.1322884	.2184627	0	1.25
Birthyear	0				
Age	1008	26.54464	3.646766	17	40
PrevCAPS	965	26.32642	23.44447	0	124
GoalsPrior	615	3.315447	6.630401	0	83
Position	0				
Substitute	1014	.2455621	.4306324	0	1
OppGK	0				
GKCaps	1014	68.03945	39.37273	2	176
CareerGoals	1014	10.07199	12.60892	0	109
Captain	1014	.0453649	.2082058	0	1
CareerCAPS	1014	57.43688	33.69422	1	184
Goalspergame	1014	.1552133	.1579583	0	.8666667
Rank	0				
TourneyRank	1014	305.1538	121.704	1	386
Experience~k	0				
GPGRank	0				
defender	1014	.2672584	.4427467	0	1
midfielder	1014	.2573964	.4374154	0	1
forward	1014	.2386588	.4264741	0	1
goalkeeper	1014	.0167653	.1284542	0	1
dmid	1014	.081854	.2742775	0	1
amid	1014	.1380671	.3451406	0	1
highstress	1014	.270217	.4442909	0	1
lowstress	1014	.729783	.4442909	0	1
GOALSxCAPS	1013	798.5301	1305.196	0	16241
Top5League~b	0				
PlayerID	0				
TeamID	0				
FirstTeam	1010	.519802	.4998552	0	1
SecondTeam	1014	.4822485	.4999314	0	1
AgeSqr	1008	717.9038	197.2505	289	1600
ScoredInTo~y	1014	.3796844	.4855478	0	1
WorldCup	1014	.2011834	.4010827	0	1
AgeCaps	963	747.1132	730.5319	0	4092
CONCACAF	1014	.0966469	.2956222	0	1
Africa	1014	.2475345	.4317928	0	1
Asia	1014	.0769231	.2666008	0	1
Copa	1014	.1706114	.3763547	0	1
Confed	1014	.0522682	.2226773	0	1
Euro	1014	.1548323	.3619233	0	1
LowPrevCAPS	1014	.1676529	.3737421	0	1
HighPrevCAPS	1014	.2061144	.4047132	0	1
MidPrevCAPS	1014	.6262327	.4840417	0	1

Table II – Breakdowns of Scenario by Shot Number

```
. tabulate Shot if highstress==1 in 1/485
```

Shot #	Freq.	Percent	Cum.
3	16	12.12	12.12
4	47	35.61	47.73
5	26	19.70	67.42
6	12	9.09	76.52
7	12	9.09	85.61
8	7	5.30	90.91
9	5	3.79	94.70
10	3	2.27	96.97
11	3	2.27	99.24
12	1	0.76	100.00
Total	132	100.00	

```
. tabulate Shot if highstress==1 in 486/1014
```

Shot #	Freq.	Percent	Cum.
3	3	2.11	2.11
4	31	21.83	23.94
5	50	35.21	59.15
6	23	16.20	75.35
7	12	8.45	83.80
8	8	5.63	89.44
9	6	4.23	93.66
10	4	2.82	96.48
11	3	2.11	98.59
12	2	1.41	100.00
Total	142	100.00	

Table III – Breakdowns of Placement of Above-Average and Below-Average Shooters

```
. tabulate Shot if LowPrevCAPS==1
```

Shot #	Freq.	Percent	Cum.
1	18	10.59	10.59
2	33	19.41	30.00
3	33	19.41	49.41
4	36	21.18	70.59
5	20	11.76	82.35
6	16	9.41	91.76
7	7	4.12	95.88
8	2	1.18	97.06
9	2	1.18	98.24
10	2	1.18	99.41
12	1	0.59	100.00
Total	170	100.00	

```
. tabulate Shot in 806/965
```

Shot #	Freq.	Percent	Cum.
1	44	27.50	27.50
2	37	23.12	50.62
3	32	20.00	70.62
4	15	9.38	80.00
5	19	11.88	91.88
6	6	3.75	95.62
7	1	0.62	96.25
8	4	2.50	98.75
11	1	0.62	99.38
12	1	0.62	100.00
Total	160	100.00	

Table IV – Correlation Tables

```
. corr Age PrevCAPS
(obs=963)
```

	Age	PrevCAPS				
Age	1.0000					
PrevCAPS	0.5430	1.0000				
dmid	-1.0000	-1.0000	-1.0000	-1.0000	1.0000	
amid	-1.0000	-1.0000	-1.0000	-1.0000	-1.0000	1.0000

```
. corr Shot highstress
(obs=1014)
```

	Shot	highst~s
Shot	1.0000	
highstress	0.6432	1.0000

Table V – 3x3 Simulations

Ex. 1 (Comp. adv. in stress state) (.9,.8,.7; .8, .7, .6)
Values are team 1's WP, team 2's WP

team 1 (vertical); team 2 (horizontal)	BP 1st, 2BP 2nd (123)	BP 1st, 2BP 3rd (132)	BP 2nd, 2BP 1st (213)	BP 2nd, 2BP 3rd (312)	BP 3rd, 2BP 1st (231)	BP 3rd, 2BP 2nd (321)
BP 1st, 2BP 2nd	.35286 / .319756	.360962 / .31706	**.351728 / .320504**	.358734/ .318884	.366736 / .31548	.365472 / .317704
BP 1st, 2BP 3rd	.350968 / .32505	.359078 / .323232	.350264 / .325064	.357492 / .323384	.365484 / .321632	.364602 / .321755
BP 2nd, 2BP 1st	.352428 / .322024	.360476 / .319328	.351536 / .322232	.358608 / .320032	.366448 / .317208	.365472 / .317704
BP 2nd, 2BP 3rd	.34996 / .330147	.357944 / .328344	.349816 / .329016	.357198 / .326016	.364812 / .325584	.36421 / .324387
BP 3rd, 2BP 1st	.34902 / .33154	.35708 / .330456	.34888 / .330552	.35634 / .328312	.36416 / .328456	.36356 / .3273
BP 3rd, 2BP 2nd	.348444 / .334384	.356432 / .333	.348624 / .332776	.356172 / .329796	.363776 / .33068	.363336 / .328784

(Bigger Comp. adv. in stress state) (.9,.8,.7; .8, .6, .4)

team 1 (vertical); team 2 (horizontal)	BP 1st, 2BP 2nd (123)	BP 1st, 2BP 3rd (132)	BP 2nd, 2BP 1st (213)	BP 2nd, 2BP 3rd (312)	BP 3rd, 2BP 1st (231)	BP 3rd, 2BP 2nd (321)
BP 1st, 2BP 2nd	.412 / .377	.452 / .336	.366 / .406	.358 / .404	.441 / .331	.393 / .369
BP 1st, 2BP 3rd	.431 / .354	.455 / .330	.397 / .370	.376 / .367	.443 / .324	.398 / .345
BP 2nd, 2BP 1st	.412 / .377	.452 / .336	.366 / .406	.358 / .404	.441 / .331	.393 / .369
BP 2nd, 2BP 3rd	.431 / .354	.455 / .330	.397 / .370	.376 / .367	.443 / .324	.398 / .345
BP 3rd, 2BP 1st	.447 / .335	.458 / .324	**.423 / .339**	**.392 / .334**	.444 / .318	.402 / .324
BP 3rd, 2BP 2nd	.447 / .335	.458 / .324	**.423 / .339**	**.392 / .334**	.444 / .318	.402 / .324

Team 1: 123 = 213 << 132 = 312 << 231 = 321
Team 2: 213 / 312 dominant strategies

EX 2: Players with comparative advantage only in general state
.9 .8 .7 becomes .5.5.5

team 1 (vertical); team 2 (horizontal)	BP 1st, 2BP 2nd (123)	BP 1st, 2BP 3rd (132)	BP 2nd, 2BP 1st (213)	BP 2nd, 2BP 3rd (312)	BP 3rd, 2BP 1st (231)	BP 3rd, 2BP 2nd (321)
BP 1st, 2BP 2nd	.448 / .318	**.453 / .335**	.483 / .300	.524 / .293	.493 / .331	.527 / .312
BP 1st, 2BP 3rd	.420 / .367	.427 / .382	.455 / .349	.498 / .341	.468 / .378	.503 / .355
BP 2nd, 2BP 1st	.446 / .328	.450 / .345	.482 / .307	.523 / .298	.490 / .339	.527 / .312
BP 2nd, 2BP 3rd	.415 / .387	.421 / .402	.453 / .364	.496 / .351	.462 / .393	.500 / .365
BP 3rd, 2BP 1st	.395 / .416	.403 / .430	.431 / .398	.474 / .388	.445 / .424	.480 / .401
BP 3rd, 2BP 2nd	.393 / .426	.400 / .440	.430 / .405	.473 / .393	.442 / .431	.478 / .406

1st lineup dominates (123 for 1st team; 213 very close)
2nd lineup for 2nd team (132 for second team)
Each lineup strictly dominates another

EX3: Proportional Decline in Abilities under stress
.9.8.7 becomes .81.72.63

team 1 (vertical); team 2 (horizontal)	BP 1st, 2BP 2nd (123)	BP 1st, 2BP 3rd (132)	BP 2nd, 2BP 1st (213)	BP 2nd, 2BP 3rd (312)	BP 3rd, 2BP 1st (231)	BP 3rd, 2BP 2nd (321)
BP 1st, 2BP 2nd	**.339 / .310**	.345 / .310	.342 / .309	.351 / .308	.353 / .309	.356 / .309
BP 1st, 2BP 3rd	.336 / .318	.342 / .318	.339 / .317	.349 / .316	.351 / .317	.354 / .316
BP 2nd, 2BP 1st	.339 / .313	.344 / .313	.342 / .311	.351 / .309	.352 / .311	.356 / .309
BP 2nd, 2BP 3rd	.335 / .324	.341 / .324	.338 / .321	.349 / .319	.350 / .321	.354 / .319
BP 3rd, 2BP 1st	.333 / .327	.340 / .327	.336 / .326	.347 / .324	.349 / .326	.353 / .324
BP 3rd, 2BP 2nd	.332 / .330	.339 / .330	.336 / .328	.346 / .326	.349 / .328	.353 / .326

Strategies identical for second team (123=132 >> 213=231 >> 312=321)
For 1st team (123=213 >> 132=312 >> 231=321)
Equilibrium is 123/123

Table VI – Shot Tabulations for Best Shooters

. tabulate Shot in 1/205

Shot #	Freq.	Percent	Cum.
1	60	29.27	29.27
2	38	18.54	47.80
3	40	19.51	67.32
4	28	13.66	80.98
5	24	11.71	92.68
6	5	2.44	95.12
7	3	1.46	96.59
8	3	1.46	98.05
9	1	0.49	98.54
11	1	0.49	99.02
12	2	0.98	100.00
Total	205	100.00	

. tabulate Shot in 1/193

Shot #	Freq.	Percent	Cum.
1	62	32.12	32.12
2	35	18.13	50.26
3	37	19.17	69.43
4	25	12.95	82.38
5	23	11.92	94.30
6	4	2.07	96.37
7	1	0.52	96.89
8	3	1.55	98.45
9	1	0.52	98.96
11	2	1.04	100.00
Total	193	100.00	

Table VII – Conversion Rates By Age Clusters

	Low-Stress Conversion %	High-Stress Conversion %
Young (1)	84.78	81.40
Average (2)	75.28	67.93
Old (3)	74.31	70.21

Table VIII - Conversion Rate By Tournament

Tournament	Overall Conversion Rate	Years Of Data
Confederations Cup / King Fahd Cup	83.02%	1995-2013
Euro Cup	80.25%	1976-2012
Copa America	79.77%	1993-2011
African Cup	76.10%	1988-2013
World Cup	70.59%	1982-2010
CONCACAF	68.37%	1991-2011
Asian Cup	61.54%	1996-2011

IX. References

- Baumann, F., Friehe, T., & Wedow, M. (2011). General ability and specialization: Evidence from penalty kicks in soccer. Journal of Sports Economics, 12(1), 81105.
- Billsberry, J. and Nelson, P. (2007) Alternatives to the penalty shootout. Paper presented at the 6th World Congress on Science and Football, Antalya, Turkey. Abstract published in the *Journal of Sports Science and Medicine, 6* (10), 96.
- Billsberry, J., Nelson, P., van Meurs, N. and Edwards, G. (2007) Are penalty shootouts racist? Paper presented at the 6th World Congress on Science and Football, Antalya, Turkey. Abstract published in the *Journal of Sports Science and Medicine, 6* (10), 98.
- Bukiet, Bruce, and Elliotte Harold. "A Markov Chain Approach to Baseball." *Operations Research* 45.1 (1997): 14-23. *JSTOR*. JSTOR, 08 Dec. 2008. Web. 05 Sept. 2013.
- Chiappori, P.-A., S. Levitt, and T. Groseclose. "Testing Mixed-Strategy Equilibria When Players Are Heterogeneous: The Case of Penalty Kicks in Soccer." *American Economic Review* 92.4 (2002): 1138-151. *JSTOR*. JSTOR. Web. 10 Sept. 2013.
- Dorhauer, Adam. “On Runs, Wins, and Two Types of Leverage Index.” *Baseball Prospectus*. (2012). Web. 5 Dec. 2013. Retrieved from http://www.baseballprospectus.com/article.php?articleid=16212
- Gibson, O. (2009, December 3). Prize money increased 61% for 2010 world cup. *The Guardian*. Retrieved from http://www.theguardian.com/football/2009/dec/03/world-cup-2010-prize-money
- Grayson, J. W. (2012, June 12). *Penalties: Are any players good at taking them? How about saving them?* Retrieved from http://jameswgrayson.wordpress.com/2012/06/19/penalties-are-any-players-good-at-taking-them-how-about-saving-them/
- Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. (2006). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *JOURNAL OF SPORTS SCIENCES,25*, 121-129.
- Keh, A. (2010, May 30). A few things to think about when lining up that kick. *New York Times*. Retrieved from http://www.nytimes.com/2010/05/31/sports/soccer/31penaltykicks.html?pagewanted=all&_r=1&

- Kuper, S., & Syzmanski, S. (2009). *Soccernomics: Why England Loses, Why Spain, Germany, and Brazil Win, and Why the US, Japan, Australia, Turkey-and Even Iraq-are Destined to Become the Kings of the World's Most Popular Sport*.
- Macaree, Paul. "Cristiano Ronaldo, Paulo Bento, And The Optimal Penalty Shootout Order." Web log post. *SB Nation*. SB Nation, 28 June 2012. Web. 5 Sept. 2013.
- Macey-Dare, R. (2013). *Optimal penalty shootouts*. Unpublished manuscript, St. Cross College, Oxford, Oxford, England.
- McGarry, Tim, and Ian Franks. "On Winning the Penalty Shoot-out in Soccer." *Journal of Sports Sciences*. Journal of Sports Sciences, 2000. Web. 4 Sept. 2013.
- O'Mard, M. K. (2012, June 27). *Blame Paulo Bento for Cristiano Ronaldo Not Taking Decisive Penalty Kick at Euro 2012*. Retrieved from http://nesn.com/2012/06/blame-paulo-bento-for-cristiano-ronaldo-not-taking-decisive-penalty-kick/
- Palacios-Huerta, I. (2012), Tournaments, Fairness and the Prouhet-Thue-Morse Sequence. Economic Inquiry, 50: 848–849.
- Slater, S. (2010, May 27). *Penalty shootouts: World cup history*. Retrieved from http://www.fourfourtwo.com/news/penalty-shootouts-world-cup-history
- Tango, Tom M. "Crucial Situations." *Hardball Times*. Hardball Times, 01 May 2006. Web. 05 Sept. 2013.