International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:13, No:5, 2019

Modeling Football Penalty Shootouts: How Improving Individual Performance Affects Team Performance and the Fairness of the ABAB Sequence

Pablo Enrique Sartor Del Giudice

Abstract**—**Penalty shootouts often decide the outcome of important soccer matches. Although usually referred to as "lotteries", there is evidence that some national teams and clubs consistently perform better than others. The outcomes are therefore not explained just by mere luck, and therefore there are ways to improve the average performance of players, naturally at the expense of some sort of effort. In this article we study the payoff of player performance improvements in terms of the performance of the team as a whole. To do so we develop an analytical model with static individual performances, as well as Monte Carlo models that take into account the known influence of partial score and round number on individual performances. We find that within a range of usual values, the team performance improves above 70% faster than individual performances do. Using these models, we also estimate that the new ABBA penalty shootout ordering under test reduces almost all the known bias in favor of the first-shooting team under the current ABAB system.

Keywords**—**Football, penalty shootouts, Montecarlo simulation, ABBA.

I. Introduction

THE The penalty shootout is a mechanism to decide the outcome of soccer matches, when a winner is required and the match is drawn after ordinary time (and extra time when it applies). Shootouts are frequent and have even decided the winner of major worldwide tournaments (e.g., World Cups 1994 and 2006, Copa Libertadores 2001, Champions League 2016, Copa América 1995, World Cup U20 2013). The rules of penalty shootouts can be seen in the Laws of the Game issued by the International Football Association Board (IFAB), whose jurisdiction over the rules is recognized by the Fédération Internationale de Football Association (FIFA); see [1]. The procedure is basically as follows:

- The referee tosses a coin and the team that wins the toss chooses whether to kick first or second.
- Each team is given five penalty shoots. They are kicked alternately. Whenever a team has scored more goals than the other could score, it wins the shootout and no more kicks are taken.
- If both teams have taken five kicks and scored the same number, kicks continue to be taken alternately until one team leads after shooting the same number of kicks.

The pressure born by players at the time of shooting, and its effect on their performance, has been widely studied. Strategies have been suggested to increase the chances of winning; see [2]. There is a known bias related with the outcome of the coin toss. Most shootouts are won by the team that takes the first kick; according to [3] the probability that it wins is close to 60.5%. Several schemes have been proposed to mitigate this bias, e.g, [4]. In an effort to reduce it, a new shooting ordering named ABBA is under test in certain unofficial and official competitions; see [5].

Penalty shootouts are often referred to as "lotteries" in sports media by players, coaches and journalists, in light of the allegedly much higher influence of luck over skill. Yet evidence shows that certain teams or nations consistently perform above or below a 50% winning rate. Take for instance the case of Uruguayan national and club major teams. Since 2000 they have played 27 official shootouts, with 8 victories and 19 defeats; see Table I, and [6]. The "null hypothesis" of being equally skilled than the average rival can be rejected with a significance $\alpha = 3\%$ by a one-sided hypothesis test for major teams and clubs (and 6% when considering all categories). The case of the England national team is also discussed in [7]. It is therefore evident that there are factors other than luck that significantly affect the team performance. When considering spending certain efforts (e.g. specific training, psychological preparation, rival analysis) to improve individual players performance, it is natural to wonder how the latter affects the probability that the team wins a shootout; i.e., to what extent an improvement of the first translates into an improvement of the second. In this article we apply three models to estimate this influence. Two of them are based on the current shootout system, while the third applies to the ABBA system currently under test. We also apply these models to estimate the reduction of the "first-kicker advantage" that might be expected under the ABBA system.

A. Definitions and Acronyms

- *ABAB*: the current system of penalty shootouts in official FIFA-ruled competitions;
- *ABBA*: the new system under test since May 2017;
- *Individual Performance (IP)*: the probability that the players of a team score when shooting a penalty kick;
- *Team Performance (TP)*: the probability that a team wins a penalty shootout;
- *5-5 tournament*: the first stage of the penalty shootout where each team kick up to 5 penalties;

Pablo Enrique Sartor Del Giudice is with the Universidad de Montevideo, Uruguay (e-mail: psartor@um.edu.uy).

- *1-1 tournament*: the second stage of penalty shootouts (only played if the 5-5 tournament is drawn), where pairs of penalties are shot until one team scores and the other one misses.

II. Analytical Model for Uniform Individual Performances

Let us start with an analytical model in which all players of a given team share the same probability of scoring when shooting a penalty kick. Let p be A's individual performance, i.e., the probability that any player of team A scores. Let q be B's individual performance. Let A_{W5} be the event "A beats B in the base 5-5 tournament", A_{D5} be the event "A and B draw the base 5-5 tournament", and A_{W1} be the event "A beats B in a 1-1 tournament". The probability $P(A)$ that A wins the penalty shootout is then

$$P(A) = P(A_{W5}) + P(A_{D5})P(A_{W1}). \tag{1}$$

$P(A_{W5})$ is the probability that a binomial random variable with parameters 5 and p is greater than another binomial random variable with parameters 5 and q, i.e.,

$$P(A_{W5}) = \sum_{i=1}^{5}\left(\binom{5}{i}p^i(1-p)^{5-i}\sum_{j=0}^{i-1}\left(\binom{5}{j}q^j(1-q)^{5-j}\right)\right) \tag{2}$$

which is a polynomial of degree 10 in p and q.

$P(A_{D5})$ is the probability that both random variables are equal, i.e.,

$$P(A_{D5}) = \sum_{i=0}^{5}\left(\binom{5}{i}^2 p^i(1-p)^{5-i}q^i(1-q)^{5-i}\right) \tag{3}$$

which is another polynomial of degree 10 in p and q.

$P(A_{W1})$ is the probability that A and B get the same outcome (both score or fail) zero or more times followed by A scoring and B missing the last 1-1 round; its probability is then

$$P(A_{W1}) = p(1-q)\sum_{i=0}^{\infty}(pq+(1-p)(1-q))^i = \frac{p(1-q)}{-2pq+p+q} \tag{4}$$

Finally, through a symbolic mathematics software, we get from (1) that

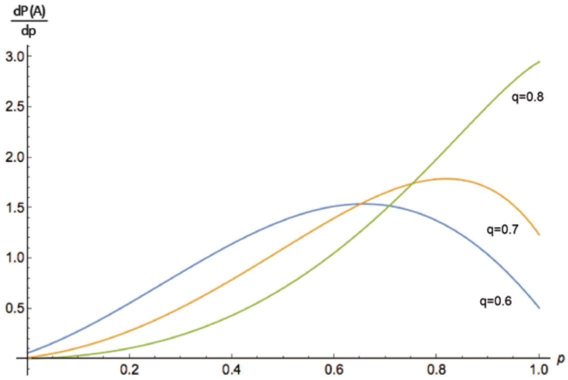

Fig. 1 Sensitivity of TP to IP as a function of the latter

$$\begin{aligned}P(A) = {} & p^5\Big(5q^4(1-q)+10q^3(1-q)^2+10q^2(1-q)^3 \\ & +(1-q)^5+5q(1-q)^4\Big)+5(1-p)p^4\Big(10q^3(1-q)^2 \\ & +10q^2(1-q)^3+(1-q)^5+5q(1-q)^4\Big) \\ & +10(1-p)^2p^3\Big(10q^2(1-q)^3+(1-q)^5+5q(1-q)^4\Big) \\ & +10(1-p)^3p^2\Big((1-q)^5+5q(1-q)^4\Big) \\ & +5(1-p)^4p(1-q)^5 \\ & +\frac{p(1-q)}{-2pq+p+q}\Big(p^5q^5+25(1-p)p^4(1-q)q^4 \\ & +100(1-p)^2p^3(1-q)^2q^3+100(1-p)^3p^2q^2(1-q)^3 \\ & +(1-p)^5(1-q)^5+25(1-p)^4pq(1-q)^4\Big)\end{aligned} \tag{5}$$

By deriving $P(A)$ we get the pace at which the TP of A improves relative to its IP improvement. Fig. 1 plots this pace in the Y-axis vs the IP (X-axis). Typical average IP are close to 0.7; for instance, the sample used by [3] has an average of 0.731. As Fig. 1 shows, the derivative is rather flat for p and q close to 0.7. For the average IP $p = q = 0.731$ we get a slope of 1.72. This suggests that the probability that A wins the shootout improves 72% faster than the IP of A does.

III. Monte Carlo Models

In this section we build Monte Carlo models for the penalty shootout that take into account how the outcome of the penalties already shot affect the IP of the next shooter. We start by computing the required sample size for a certain confidence interval. Then we develop a model for the current system (ABAB) and a model for the one that is being tested (ABBA). We apply them to evaluate the TP improvement and compare both models' 'first-shooter' biases.

A. Crude Monte Carlo

Let $\Phi \in \mathbb{R}$ be a random variable. Monte Carlo simulates a sample $\Phi_1, \ldots, \Phi_N$ of N independent random variables with

TABLE I
PERFORMANCE OF URUGUAYAN NATIONAL TEAMS AND CLUBS IN OFFICIAL SHOOTOUTS SINCE 2000

Year	Category	Competition	Uruguayan Club/Team	Stage	Rival	Outcome
2000	Major	Copa Libertadores	Peñarol	8th-final	Palmeiras	Lost
2000	Major	Copa Libertadores	Nacional	8th-final	Bolivar	Lost
2001	Major	Copa América	Uruguay	3rd place	Honduras	Lost
2002	Major	Copa Libertadores	Peñarol	Quarterfinal	Sao Caetano	Lost
2002	Major	Copa Sudamericana	Nacional	Semifinal	Atlético Nacional	Lost
2003	Major	Copa Libertadores	Nacional	8th-final	Santos	Lost
2003	Major	Copa Sudamericana	Nacional	8th-final	Libertad	Lost
2004	Major	Copa América	Uruguay	Semifinal	Brasil	Lost
2004	U16	Sudamericano	Uruguay	Quarterfinal	Brasil	Won
2005	Major	WC Qualif. Playoff	Uruguay	Single	Australia	Lost
2005	Major	Copa Sudamericana	Nacional	8th-final	Boca Juniors	Won
2006	Major	Copa Sudamericana	Danubio	32nd-final	Tacuary	Won
2007	Major	Copa Libertadores	Defensor Sporting	Quarterfinal	Gremio	Lost
2007	Major	Copa América	Uruguay	Semifinal	Brasil	Lost
2009	Major	Copa Sudamericana	River Plate	Quarterfinal	San Lorenzo	Won
2009	U17	World Cup	Uruguay	Quarterfinal	España	Lost
2010	Major	World Cup	Uruguay	Quarterfinal	Ghana	Won
2011	Major	Copa América	Uruguay	Quarterfinal	Argentina	Won
2012	U20	Libertadores U20	Defensor Sporting	Quarterfinal	Universitario de Deportes	Won
2012	U20	Libertadores U20	Defensor Sporting	Semifinal	Union Española S.A.D.P	Won
2013	Major	Copa Libertadores	Nacional	8th-final	Real Garcilaso	Lost
2013	U20	World Cup	Uruguay	Semifinal	Irak	Won
2013	U20	World Cup	Uruguay	Final	Francia	Lost
2013	Major	Copa Confederaciones	Uruguay	3rd place	Italia	Lost
2014	Major	Copa Libertadores	Defensor Sporting	8th-final	The Strongest	Won
2014	Major	Copa Sudamericana	Peñarol	8th-final	Estudiantes	Lost
2015	Major	Copa Sudamericana	Juventud	16th-final	Emelec	Lost
2015	Major	Copa Sudamericana	Defensor Sporting	8th-final	Lanús	Won
2015	U15	Sudamericano	Uruguay	Final	Brasil	Lost
2015	U20	World Cup	Uruguay	8th-final	Brasil	Lost
2016	U20	Libertadores U20	Liverpool	Semifinal	Deportivo Tuluá	Won
2016	Major	Copa Libertadores	Nacional	Quarterfinal	Boca Juniors	Lost
2016	Major	Copa Sudamericana	Wanderers	32nd-final	O'Higgins	Won
2016	Major	Copa Sudamericana	Plaza Colonia	32nd-final	Blooming	Lost
2016	Major	Copa Sudamericana	Wanderers	8th-final	Atlético Junior	Lost
2017	Major	Copa Sudamericana	Danubio	16th-final	Sport Recife	Lost
2017	U20	World Cup	Uruguay	Quarterfinal	Portugal	Won
2017	U20	World Cup	Uruguay	Semifinal	Venezuela	Lost
2017	U20	World Cup	Uruguay	3rd place	Italia	Lost

the same probability distribution as Φ, and proposes the estimator $\overline{\Phi_N}$:

$$\overline{\Phi_N} = \frac{1}{N}\sum_{i=1}^{N}\Phi_i. \tag{6}$$

As corollary from Kolmogorov's Strong Law, the estimator $\overline{\Phi_N}$ is consistent for the expected value $E(\Phi)$. Moreover, as $E(\overline{\Phi_N}) = E(\Phi)$ for all $N \geq 1$, it is also unbiased. Let $\Phi \in \{0,1\}$ so that $\Phi = 1$ when team A wins and $\Phi = 0$ when B wins. We want to estimate the TP of A: $P(\Phi = 1) = E(\Phi)$ using a Monte Carlo simulation. If we want to constrain the radius to 0.0001, i.e. 0.01%, with a confidence level of 95%, we need a sample size of

$$N = \left(\frac{1.96}{2\times 0.0001}\right)^2 \approx 10^8 \tag{7}$$

B. Monte Carlo Model for the Current System ABAB

In our first Monte Carlo model each trial $\Phi_1, \ldots, \Phi_N$ corresponds to the outcome of a penalty shootout, with $\Phi_i = 1$ if A wins and $\Phi_i = 0$ if B wins for $i = 1 \ldots N$. To build each trial, penalties are alternatively simulated for A and B by drawing Bernoulli's variables, where 1 means "scored" and 0 means "missed", until one of both teams wins the shootout. The probability that a "1" is drawn for a certain penalty depends on three factors:

- whether the shooter's team is lagging, even or ahead in the score at the time of shooting. We represent this by means of a variable taking the values $\{-1,0,1\}$;
- the round number, i.e., the number of penalties kicked by the team once this penalty is kicked (we represent this by means of a variable taking the values $\{1,2,3,4,5,6\}$ where 6 is used for any penalty during the 1-1 tournament);
- whether the player shoots first or second in that round (we represent this by means of a binary $\{0,1\}$ variable).

To model the probability of scoring we use the database shown in Table II, which reproduces Table VI of [3]. This databse was built upon 1338 penalties held during shootouts in major soccer competitions around the world. It includes the scoring rates according to the three factors above as well as the sample size for each combination. Small number effects are present; observe, for instance, that all penalties shot in the second round by the first shooter trying to tie the game were scored. According to Table II this corresponds to just 16 penalties, all successful. Hence we built a logit regression on top of the 1338 instances with the three variables above

International Journal of Medical, Medicine and Health Sciences
ISSN: 2517-9969
Vol:13, No:5, 2019

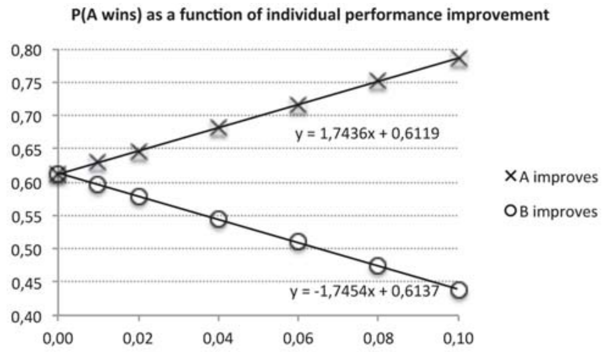

Fig. 2 TP of A vs IP increase (using the current system ABAB)

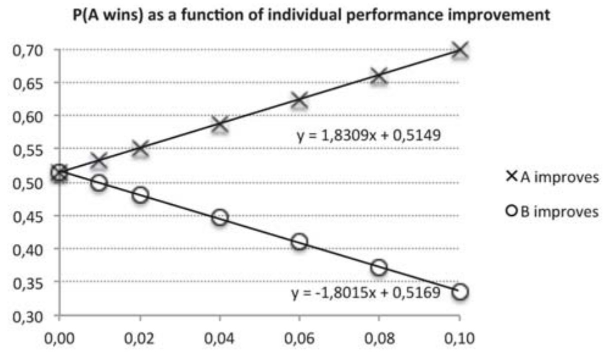

Fig. 3 TP of A vs IP increase (using the proposed system ABBA)

mentioned as the independent ones. By doing so we got the probabilities shown in Table III with a p-value of 0.0016.

Since we are simulating the ABAB system, A always plays the role of "1st. shooter" whilst B plays that of "2nd. shooter". The probability $P(A)$ that A wins the shootout can thus be estimated by $\overline{\Phi_N} = \frac{1}{N}\sum_{i=1}^{N}\Phi_i$. With a sample size of $N = 10^8$ we got $\overline{\Phi_N} = 0.6117$. The confidence interval is approximately that of Section III-A, i.e., a radius of 0.0001 with a confidence level of 95%. This result is pretty close to 60.5%, the success rate of the sample for the first-shooting team, which suggests that a model based on the two factors above mentioned performs very well[1].

It is of our interest to analyze the influence of IP on the TP. To do so we ran two sets of simulations. The first set includes six simulations where the IP of A given by Table III were altered by adding a probability of $1\%, 2\%, 4\%, 6\%, 8\%$ and 10% respectively[2]. The second set includes six simulations where the IP of B were altered the same way. Table IV shows the estimators of $P(A)$ returned by each simulation. As Fig. 2 shows, the TP of A fit very well two lines with slopes 1.74 (when A improves) and -1.75 (when B improves). This suggests that, over the range of IP improvement studied, the TP of A grows approximately by 1.74% for every percentual point of individual performance gained by A, whilst it diminishes by 1.75% per every percentual point of individual performance gained by B.

C. Monte Carlo Model for the Alternative System ABBA

Our second Monte Carlo model implements the "ABBA" ordering followed in the alternative shootout system under test. As in the model of Section III-B, the IP are taken from Table III to draw the outcome of each penalty kick and therefore the outcome of the shootout. In the first round (the first two penalties of the shootout) team A plays the role of "1st. shooter" and team B is the "2nd. shooter" (the IP is taken from the columns labeled with the role played). This applies also to rounds 3 and 5. In rounds 2 and 4, team A plays the role of "2nd. shooter" while team B is the "1st. shooter". If there is no winner after the 5-5 tournament, the 1-1 tournament proceeds starting with A as "2nd. shooter" and B as "1st. shooter" and switching these roles after each round until a winner emerges. As in Section III-B, the IP is determined by the score state (lagging, even or ahead), the round number, and the role played. We estimated $P(A)$ running this model with $N = 10^8$ and got $\overline{\Phi_N} = 0.5157$ with radius 0.0001 and confidence level of 95%. Observe that there is still a bias in favor of team A, although much lower than in the model for the current system (1.57% instead of 11.17%, i.e., the proposed system corrects an 86% of the bias evidenced by the current system). This result is similar to the one obtained by [8] using an analytical model for 1-1 tournament shootouts.

Once again we analyze the influence of IP on the team performance by running two sets of simulations analogous to those of Section III-B. Table V shows the estimators of $P(A)$ returned by each simulation. Again, as Fig. 3 shows, the A's TP fit very well two lines with slopes 1.83 (when A improves) and -1.80 (when B improves).

IV. Conclusions

In this article we introduced three models to study the probability that a team wins a penalty shootout and its sensitivity to improvements in the IP of the team's players:

- a closed expression for a model where both teams have uniform IP p and q respectively;
- a Monte Carlo simulation model for the current shootout system (ABAB), which takes into account the round number, the score state (lagging, even or ahead), and the kicking order within the round;
- an analogous Monte Carlo simulation model for the new proposed system (ABBA) under test.

Our tests with the Monte Carlo models show that when a team improves its IP by x, its TP increases 1.74 (ABAB) and 1.83 (ABBA) times x. Hence we quantified this "magnifying" effect, that could be expected in light of the repetitive nature of the shootouts (since every team kicks at least three penalties). This could be summarized as follows: "the TP improves 74% or 83% (ABAB or ABBA) faster than the IP does". As an example, consider the case of the Uruguayan club and national teams since 2000. The deficit of team performance relative

[1] Other factors as the difference in score at the time of shooting might account for the difference of 0.67%.

[2] The altered individual performances where set to 100% whenever the value taken from the table plus the added probability exceed 100%.

TABLE II
SCORING PROBABILITIES AND WINNING FREQUENCIES BY TEAM, ROUND AND PARTIAL SCORE

	1st Team			2nd Team		
	Behind	Even	Ahead	Behind	Even	Ahead
Round 1:						
Scoring Probability	-	78.9	-	75.2	59.3	-
Percent Win Shoot-out	-	60.2	-	32.7	66.7	-
Penalty Kick Importance		56.5		93.4	39.3	
N	-	128	-	101	27	-
Round 2:						
Scoring Probability	100	74.7	96.0	82.2	65.8	-
Percent Win Shoot-out	31.3	57.5	88.0	32.2	57.9	-
Penalty Kick Importance	-	32.2	30.6	62.7	61.2	
N	16	87	25	90	38	0
Round 3:						
Scoring Probability	80.0	76.8	76.5	63.2	69.4	40.0
Percent Win Shoot-out	24.0	59.4	88.2	23.0	72.2	100
Penalty Kick Importance	115.7	67.3	21.9	62.7	66.4	14.3
N	25	69	34	87	36	5
Round 4:						
Scoring Probability	76.7	71.7	75.0	66.2	69.4	77.8
Percent Win Shoot-out	13.3	62.3	88.6	21.1	75.0	100
Penalty Kick Importance	150.0	68.2	19.8	125.1	50.1	14.8
N	30	53	44	71	36	9
Round 5:						
Scoring Probability	74.1	76.2	71.4	62.5	70.0	-
Percent Win Shoot-out	14.8	52.4	96.4	30.0	83.3	-
Penalty Kick Importance	112.5	101.8	31.1	156.9	63.5	
N	27	42	28	40	30	-
Round 6+:						
Scoring Probability	-	67.5	-	68.5	65.4	-
Percent Win Shoot-out	-	58.8	-	24.1	76.9	-
Penalty Kick Importance		90.0		153.5	84.7	
N	-	80	-	54	26	-

TABLE III
IP OUT OF A LOGIT REGRESSION MODEL ON SHOOTING ORDER, ROUND AND PARTIAL SCORE

Round	1st. Team			2nd. Team		
nbr.	behind	even	ahead	behind	even	ahead
1		80.3		74.5	72.9	
2	80.0	78.7	77.2	72.6	70.9	69.1
3	78.4	77.0	75.4	70.6	68.8	67.0
4	76.7	75.2	73.6	68.5	66.7	64.8
5	74.9	73.3	71.6	66.4	64.5	
6+		71.3		64.1	62.2	

TABLE IV
TP OF A WHEN INCREASING THE IP OF A AND B (USING THE CURRENT SYSTEM ABAB)

Increase of Individual Performance	Applied to A	Applied to B
0.00	0.6117	0.6117
0.01	0.6292	0.5958
0.02	0.6467	0.5794
0.04	0.6819	0.5457
0.06	0.7168	0.5108
0.08	0.7515	0.4745
0.10	0.7859	0.4370

TABLE V
TP OF A WHEN INCREASING THE IP OF A AND B (USING THE PROPOSED SYSTEM ABBA)

Increase of IP	Applied to A	Applied to B
0.00	0.5157	0.5157
0.01	0.5333	0.4986
0.02	0.5512	0.4812
0.04	0.5873	0.4458
0.06	0.6241	0.4097
0.08	0.6613	0.3729
0.10	0.6988	0.3355

to their rivals, close to 17%, might be eliminated with an improvement of individual performance of just about 10%.

The database of shootouts used in [3] results in a TP of 60.5% for the team that kicks first. Table 6 in the same article presents the corresponding IP totaled by three factors: round number, score state and shooting order. Our Monte Carlo model, based exclusively on a logit regression model built on their sample returns a TP of 61.17% for the team that kicks first. This suggest that the three factors are enough to model quite well the outcome of the shootouts.

Finally, according to our simulation models, under the proposed ABBA system, the team that kicks first has a TP of 51.57%. The bias in its favor is thus reduced from 11.17% to 1.57% (an 86% reduction). The proposed system mitigates almost all the influence of being the first or second shooter, although it does not completely eliminate the bias in favor of the first one. It still remains to be seen whether the FIFA considers that this bias reduction justifies changing the game rules.

References

[1] IFAB. Laws of the game - 2016/17. http://www.theifab.com, June 2016.
[2] Tim McGarry and Ian M. Franks. On winning the penalty shoot-out in soccer. *Journal of Sports Sciences*, 18(6):401–409, 2000. PMID: 10902675.
[3] Jose Apesteguia and Ignacio Palacios-Huerta. Psychological pressure in competitive environments: Evidence from a randomized natural experiment. *American Economic Review*, 100(5):2548–64, December 2010.
[4] Steven J. Brams and Mehmet S. Ismail. Making the rules of sports fairer. *SSRN Electronic Journal*, 2016.
[5] UEFA. Penalty shoot-out trial at UEFA final tournaments, May 2017.
[6] Pablo Regent and Pablo Sartor. La cultura como desventaja deportiva (Culture as a competitive disadvantage). *Revista de Negocios IEEM*, XX(3):34–45, June 2017.
[7] Andrew Anthony. *On penalties*. Yellow Jersey Press, May 2001.
[8] Federico Echenique. ABAB or ABBA? The arithmetics of penalty shootouts in soccer. Technical report, California Institute of Technology, 2017.