Fooled by performance randomness: over-rewarding luck

Romain Gauriot[1*]   Lionel Page[2]

[1]University of Sydney
[2]Queensland University of Technology

Abstract

We provide evidence of a violation of the informativeness principle whereby lucky successes are overly rewarded. We isolate a quasi-experimental situation where the success of an agent is as good as random. To do so, we use high quality data on football (soccer) matches and select shots on goal which landed on the goal posts. Using non scoring shots, taken from a similar location on the pitch, as counterfactuals to scoring shots, we estimate the causal effect of a lucky success (goal) on the evaluation of the player's performance. We find clear evidence that luck is overly influencing managers' decisions and evaluators' ratings. Our results suggest that this phenomenon is likely to be widespread in economic organizations.
Keywords: contract theory, informativeness principle, quasi-experiment, outcome bias, behavioural economics.
JEL codes: D23, D83, D86, L14

1 Introduction

The informativeness principle (Hölmstrom 1979, 1982) is a corner stone of contract theory (Bolton and Dewatripont 2005). It states that all valuable signals of performance should be rewarded. A signal is valuable if it is *informative* about the agent's performance. Empirical observation reveals that this principle is often violated by real world contracts. Some informative signals of performance are often not incorporated in explicit contracts (Prendergast 1999). We investigate here another possible violation of this principle: when *uninformative signals* are actually taken into account and given weight in the evaluation of the agent.

In organizations, accurate and objective signals of performance are typically unavailable and principals have to rely on subjective performance evaluation. Unfortunately subjective evaluation is prone to biases. Managers' incentives may for instance limit the accuracy of subjective performance evaluation in organizations (Prendergast and Topel 1993, 1996, Bol 2011). The literature in psychology suggests another type of possible bias: cognitive biases. Making subjective judgements about performance is not trivial.

The relative success or failure of an agent is often driven in part by his actions and in part by random and unforeseeable circumstances. In such situations, the evaluator has the difficult task to evaluate the agents' performance based on his *ex ante* merits rather than the elements of *luck* which may have influenced his outcome.

In typical real world situations, researchers can neither access all the explicit and implicit information the evaluator has about the agent, nor identify perfectly the informativeness of naturally occurring signals of performance. As a consequence it is hard to assess whether an evaluator updates his/her information adequately given his/her observation of a signal of performance. This paper overcomes this problem by carefully isolating situations where, conditional on a specific signal of performance, agents' success can be considered as good as random and is therefore not informative.

We take advantage of the large amount of data on performance in sporting contests to isolate specific situations where a player's success can be considered as primarily driven by luck. Specifically, we look at football (soccer) players. Conditional on hitting the post while attempting to score, we show that there are no significant differences in the average performance of the players who either score or do not score after hitting a post. These selected events provide a quasi-experimental setting in which we can test whether an *uninformative* outcome (goal being scored or not) plays a role in players' evaluation.

We find that players who score are given significantly more playing time in the following match by team managers than players who do not score (in the same situation). Looking at third party expert evaluators (journalists), we find that they give ratings which are two thirds of a standard deviation higher when a goal is scored rather than when it is not. We also find a similar result in the ratings of football fans who can be considered as another type of principals as consumers of the entertainment generated by the players. Even though the shots' outcomes can be considered as good as random, these outcomes have a substantial effect on judgements and decisions related to the players' performance.

This paper contributes to the understanding of the extent to which the informativeness principle is respected in the field. It investigates a specific type of deviation generated by what has been named the "outcome bias". Following Baron and Hershey (1988), studies in psychology have suggested that individuals tend to give too much importance to information about the outcome when trying to assess the quality of the decisions made by an agent. This research has led to a concern for such a potential bias in medical decisions (Chapman and Elstein 2000), ethical judgements (Gino, Shu and Bazerman 2010), and legal decisions (Alicke, Davis and Pezzo 1994).[1] However, there is scant evidence for the existence, prevalence and magnitude of such an outcome bias in the field. Whether or not it is a relevant concern for economic organizations depends on whether such evidence exists. Overall our results indicate that this bias is likely to be widespread and be an influence on subjective performance evaluations in organizations.

The remainder of the paper is as follows. In Section 2, we present the conceptual framework underlying the informativeness principle. Section 3 presents our quasi-experimental data and empirical strategy. Section 4 presents the main results while Section 5 provides a range of robustness checks. Section 6 concludes.

---

[1]Even though no studies have provided empirical evidence about it. A similar concern has arisen in the financial industry where high performers may be drawn from a sample of employees who have previously adopted highly risky strategies (Coates 2012, p. 259).

2 Conceptual framework

Let's consider a situation where a principal wants to assess the quality of the actions (decisions/effort) made by an agent. We follow here the formal framework of Bolton and Dewatripont (2005, p. 130) which is an adaptation of Hölmstrom (1979). Suppose that the outcome q following the agent's actions is binary with $q = 1$ describing a success and $q = 0$ a failure. The agent has to choose an action $a \in \mathbb{R}$ which influences the probability of success $\mathbb{P}(q = 1|a) = p(a)$. This action is not observed by the principal. However the principal also observes a binary signal $s \in \{0, 1\}$ which is influenced by the agent's action such that $\mathbb{P}(q = i, s = j|a) = p_{ij}(a)$.

Suppose that the principal's utility function is

$$V(q - w)$$

and that the agent's utility function is a function of the incentives provided by the principal, w and the action a which is costly:

$$u(w) - a$$

The principal aims to use the information on the outcome q and the signal s in the best possible way to assess the quality of the agent's action a and reward it accordingly to provide the agent with the right incentives. When both parties are risk averse, the agent's incentives w_{ij} are conditioned on the values of $q = i$ and $s = j$. Specifically the optimal incentives w_{ij} respect the condition:

$$\frac{V'(i - w_{ij})}{u'(w_{ij})} = \lambda + \mu \frac{p'_{ij}(a)}{p_{ij}(a)}$$

The informativeness principle imposes that q and s should impact the agents' incentives in so far as they are informative about the unobserved action a. While the informativeness principle is usually applied to the signal s, it also applies to the outcome q. It is clear from this framework that "the outcome q can be used as a signal about the action which is not directly observed" (Hölmstrom 1979).

An outcome bias can be thought as the situation when an evaluator overweights the outcome q relative to its informational content. It leads him to overestimate the ex-ante performance of the agent after a success and underestimate the performance of the agent after a failure. Let $\widehat{a}(q)$ be the estimate of the agent's action a by an evaluator and $\widetilde{a}(q)$ be the Bayesian posterior mean from a rational evaluator.

Definition 1 (Outcome bias) *An evaluator is characterised by an outcome bias when evaluating an agent if:*
$$\begin{cases} \widehat{a}(q) - \widetilde{a}(q) > 0 & if \quad q = 1 \\ \widehat{a}(q) - \widetilde{a}(q) < 0 & if \quad q = 0 \end{cases}$$

In practice, it is difficult to assess whether such a bias exists in the field. The rational belief $\widetilde{a}(q)$ is unobservable because neither the prior of the evaluator, nor the precise informational content of the outcome q are observed. It is therefore not possible to determine whether the outcome q is used appropriately such that $\widehat{a}(q) = \widetilde{a}(q)$, or not. The empirical observation of a positive correlation between the outcome q and the evaluations $\widehat{a}(q)$ may simply reflect that evaluators use their information appropriately. For this

3

reason, evidence of outcome bias in the field is only tentative (such as surveys of real evaluators facing hypothetical scenarios).

Note however that, for a given signal $s = j$, the outcome may be uninformative:

$$p'_{ij}(a) = p'_{i'j}(a) = 0, \qquad (1)$$

Conditional on the signal s, the ex-ante effect of a on the probability of success is the same whether the ex-post outcome is a success or not. In such a situation nothing about the action a can be inferred from the outcome.[2] The signal s is a sufficient statistic for the pair (s, q) with respect to a. The information on the outcome should not play a role in the evaluation of the agent. An outcome bias can therefore be tested in situations where, conditional on an informative signal $s = j$, the outcome q is as good as random in that it is not correlated with unobserved actions from the agent. This fact is at the root of our empirical strategy. We carefully look for situations where the outcome of an agent's action, success or failure, can be considered as good as random.

3   Empirical strategy and data

3.1   Empirical strategy

We look here at the evaluation of players' performance in Association Football (soccer). In team sports, players' personal incentives are imperfectly aligned with their team which lead to the typical moral hazard problem in principal-agent interactions. Players may exert a suboptimal level of effort (Marburger 2003), or they may adopt suboptimal strategies to reap personal rewards (Gauriot and Page 2015). In football, insufficient effort levels may prevent players from being in the best shooting positions. And personal incentives may lead players to privilege shooting at goals rather than passing the ball. The players' effort and decision motives are imperfectly observed. To address this principal agent problem, signals of performance should be rewarded following the informativeness principle.

Using a high quality data set on football matches, we isolate events where shots aimed at the opposition goal landed "on the post". When players kick the ball from far away at high speed they are unable to perfectly control the trajectory of their shot. While most of the shots end up either in or out of the goal (sometimes far off the frame), some of the shots hit one of the posts of the goal frame. A difference of a few centimetres makes then the difference between the shot being *in* (goal) or *out* (no goal). It is reasonable to assume that the difference in average performance between scoring and non-scoring players is negligible (we provide evidence in support of this assumption). We therefore have the setting for a quasi-natural experiment whereby players with similar signals of performance (hitting the post) have a final outcome (goal/no goal) which is as good as random. In such a setting a success does not add any information about the performance relative to the observed signal.

To ensure that our identification assumption is as credible as possible, we use the information on the precise location of the shots. We are able to match each shot *in* to a shot *out* taken from a very similar location on the pitch and vice versa (within a radius of 45cm on average). Let Y be a variable of interest, such as the judgement of an evaluator about the performance of the player. We use matched observations to estimate

---

[2]This specific case had been noted by Hölmstrom (1979) himself.

the potential value of Y if the outcome of the shot had been different. Following Abadie and Imbens (2006) the matching estimator is:[3]

$$\widehat{\tau}_M = \frac{1}{N} \sum_{i=1}^{N} \left(\widehat{Y}_i(1) - \widehat{Y}_i(0) \right) \tag{2}$$

Let $\mathcal{J}_M(i)$ be the set of indices for the first M matches for observation i. We can define for any shot on post ending in a goal the potential value:

$$\widehat{Y}_i(1) = \begin{cases} Y_i & \text{if a goal is scored} \\ \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j & \text{if no goal is scored} \end{cases}$$

And for any shot on post ending in a near miss the potential value:

$$\widehat{Y}_i(0) = \begin{cases} \frac{1}{M} \sum_{j \in \mathcal{J}_M(i)} Y_j & \text{if a goal is scored} \\ Y_i & \text{if no goal is scored} \end{cases}$$

In the following empirical analysis, our default specification is a one-nearest neighbor matching using a Euclidean distance (i.e. matching each shot with the closest counterfactual on the pitch).

3.2   Data

Our dataset includes detailed information on all goals scored and all shots that hit the frame of the goal from the 2006-2007 season to the 2015-2016 season in the five major European leagues: England, France, Germany, Italy, Spain. This dataset comes from the company Opta, a private company which collects and distributes data on different sports. In particular, they collect in-play information on football matches.[4] They record information on events such as passes or shots to the goal with a time stamp and spatial coordinates of the shots. The dataset contains: the name of the player making the shots, the time at which the shots on post occurred, the spatial coordinates of where the shots were taken. We supplement this data on shots with information on the matches where they took place: goals scored, the identity of the players scoring, the times of the goals, teams' line-up, and players' market values.[5] Table 1 describes our dataset.[6] In total we have the line-ups of $18,228$ matches, in which $13,066$ shots on a post were made ($2,387$ resulted in a goal and $10,679$ bounced out of the goal). Figure 1 represents the spatial coordinates of where these shots where taken on the football field.

To investigate the existence of an outcome bias, we collected the team line-ups for all the matches in the five European leagues over the period 2006-2016. We use this information to record the managers' decision to give playing time to the players. We also collected the ratings given by professional sport journalists about the players' individual performance at the end of the match. We obtain these ratings from major newspapers

---

[3]In order to control for the possible bias stemming from the use of more than one matching dimension, we use the bias corrected estimator from Abadie and Imbens.

[4]More information on how they collect this data can be found on their website: `http://www.optasports.com/about/how-we-do-it/the-data-collection-process.aspx`

[5]We collected this information from the website Transfermarkt.

[6]Post on own goal are excluded. All shots that touched the goal post of the opposite team are included.

Competition	All shots on posts			With ratings		
	Matches	Post out	Post in	Matches	Post out	Post in
Bundesliga	3,060	1,906	425	3,055	1,778	396
Ligue 1	3,800	2,004	421	2,384	1,181	273
PL	3,800	2,330	513	1,634	1,019	227
Serie A	3,799	2,221	509	-	-	-
Liga	3,773	2,218	519	-	-	-
Total	18,232	10,679	2,387	7,073	3,978	896

Table 1: Dataset description. Excluding post on own goal.

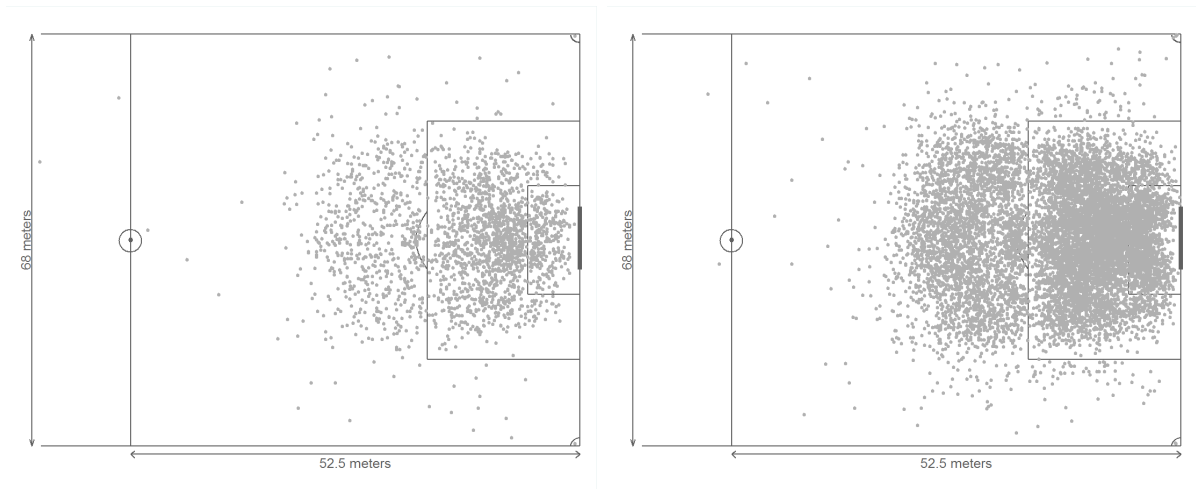Figure 1: Graphical representation of the starting point of shots ending on the posts. On the left panel the post *in* ($N = 2,387$) and on the right the post *out* ($N = 10,679$) .

and TV channel sources for three major European leagues of football: English, French and German leagues from 2006 to 2016.[7] Table 2 presents summary statistics of those ratings. For comparison purposes, we have rescaled all ratings to be from 0 to 1, where 1 is the best rating.[8] Finally, we collected ratings from football fans for a small sub-sample of observations in the English Premier League[9] for the seasons 2012-2013 and 2013-2014. This corresponds to 477 shots on posts *in* and 118 *out*.

---

[7]Ratings were collected from *Skysports* for the English Premier League for the season 2011-2012 to 2015-2016, *L'Equipe* for the French Ligue 1 for the season 2007-2008 to 2012-2013, *Kicker* for the German Bundesliga for the season 2007-2008 to 2015-2016.

[8]The rating scales are different across countries. In Germany the ratings are from 1 to 6, where 1 is the best rating. In England and France the rating are from 1 to 10, where 10 is the best rating.

[9]At the end of the match on the Skysports website, the fans are given the opportunity to rate the players.

	Average	Min	Max	SD	N
Overall	0.548	0	1	0.164	167,233
Starting player	0.539	0	1	0.189	154,036
Substitute	0.529	0	1	0.145	13,197
Forwards	0.524	0	1	0.193	33,056
Midfielder	0.549	0	1	0.166	42,438
Defender	0.534	0	1	0.153	77,586
Keeper	0.615	0	1	0.141	14,153
Home team	0.561	0	1	0.166	83,581
Away team	0.534	0	1	0.166	83,652
Winning team	0.621	0	1	0.141	61,441
Draw	0.552	0	1	0.145	43,499
Losing team	0.472	0	1	0.164	62,293
Non-scorer	0.531	0	1	0.161	151,940
Scorer	0.710	0	1	0.131	15,293
Score at least 2	0.865	0	1	0.093	1,689
Score at least 3	0.960	0.8	1	0.053	170
Score own goal	0.440	0	0.9	0.189	524
Does not hit the post	0.545	0	1	0.163	162,539
Hit the post at least once	0.636	0	1	0.163	4,694
Score once by hitting the post	0.732	0.2	1	0.133	895

Table 2: Players' ratings summary statistics.

4 Results

4.1 Effect on managers' decisions

Teams' managers are in charge of selecting the players along the season. They not only have to make judgements on the performance level of each player at a given moment in time, they have to act on it when selecting players for each match line-up. We look at three measures of the players' chances to be fielded in the next match: probability to play in the next match, probability to be in the main line-up (i.e. starting the match), number of minutes played.[10] Table 3 shows the results. The first column presents the effect on the probability to play in the match $t+1$, the second column shows the difference in the probability to start versus being on the bench, and the third column shows the difference in playing time in minutes. In addition to the spatial coordinates, we do an exact match on whether the player making the shot is a substitute or not.

When looking at all the shots on posts, scoring players see their playing time increase. After a shot on a post, a player will play 2 more minutes on average in the next match relative to a player who did not score in the same situation. Table 3 splits the observations as a function of the situations when the shot on goal was observed. An outcome bias could be expected to be more pronounced in situations where a goal would make a larger difference for the match's outcome. It is indeed what we observe. The effect on managers' decisions is larger for situations where the shot on goal can move the match from a draw to a victory (at the time of the shot). In such situations, the player benefits from 5 more

---

[10]We use the previous match $(t-1)$ as a baseline to avoid a possible effect of goal scoring on substitution decisions in t.

	ΔPlay	ΔStart	Δ Minutes
All situations			
Effect of Scoring	0.007	0.022	2.21*
p	0.530	0.115	0.041
N	$12,331$	$12,331$	$12,331$
Loss $\rightarrow$ Loss			
Effect of Scoring	-0.005	0.045	2.84
p	0.921	0.352	0.433
N	939	939	939
Loss $\rightarrow$ Draw			
Effect of Scoring	-0.021	0.010	1.25
p	0.461	0.783	0.638
N	$2,148$	$2,148$	$2,148$
Draw $\rightarrow$ Win			
Effect of Scoring	$0.035^{\dagger}$	0.060**	4.91**
p	0.083	0.007	0.004
N	$5,641$	$5,641$	$5,641$
Win $\rightarrow$ Win			
Effect of Scoring	0.014	-0.009	1.19
p	0.442	0.734	0.518
N	$3,583$	$3,583$	$3,583$
SD	0.509	0.540	43.56
N	$474,014$	$474,014$	$474,014$

Table 3: Effect of scoring (after a shot on the post) on manager's decisions. The dependent variable is the difference between the decision in match $t+1$ and in match $t-1$. One-nearest neighbor matching with Euclidean distance. For the main estimate, the average distance between the locations of a shot and its counterfactual is 45cm. The first and last match of each season are not included. * significance at the 5% level, ** significance at the 1% level.

minutes of playing time in the next match relative to the previous time. He is also 6 percentage points more likely to start the next match.

How does the effect of scoring after hitting the post compare to the effect of scoring any goal on managers' decision? Table 4 presents the raw effect on fielding decisions of a player's goal. Overall, a player is 8% more likely to start the next match as well as play 6 more minutes more on average. When looking at situations where the player landed a shot on the post, we observe that even when the player does not score he is 4% more likely to play in the next match and plays on average 2.64 minutes more. This effect reflects the fact that landing a shot on the post likely signals a good performance.

The extent to which managers overly reward luck can be seen in the difference in decisions following shots on posts.[11] This difference (5.26-2.64) represents more than

---

[11]Note that the numbers from Table 4 are just average differences. The differences in point estimates are therefore not exactly identical to our estimate of the outcome bias using a matching estimator in Table 3. The magnitude of this difference is however very close from our matching estimates.

	ΔPlay	ΔStart	ΔMinute
Any goal	0.040***	0.083**	6.05***
p	< .001	0.006	< 0.001
No goal from post	0.012**	0.044***	2.64***
p	0.009	< 0.001	< 0.001
Goal from post	0.024**	0.069***	5.26***
p	0.008	< 0.001	< 0.001
N	474,014	474,014	474,014

Table 4: Effect of scoring on the managers' decisions. Comparing the effect of any goal on managers decisions to situations where the player puts the ball on the post and either scored or not. The dependent variable is the difference between the decision in match $t + 1$ and in match $t - 1$. The first and last match of each season are not included. Std error clustered by player-season. ** significance at the 1% level, *** significance at the 0.1% level.

40% of the average effect from scoring a goal. This effect is therefore large in that regard. Our results suggest this bias may represent a substantial factor driving managers' decision to field players.

4.2 Effect on journalists' ratings

On a large subsample of the dataset, we observe the ratings given to players by sport journalists. Table 5 shows the results of our matching estimations for the effect of scoring on journalists post-match ratings. Conditional on hitting the post, scoring a goal increases the rating of a player by 0.113 ($p < 0.001$, $N = 4,874$). Moreover, the effect not only affects the players ratings but beyond him the evaluation of the whole team performance: the average rating of the team, excluding the player scoring, increases by 0.015 ($p < 0.001$, $N = 4,874$). In order to check that the identification assumption – outcomes are as good as random – is right, we look at the average ratings of these players over the previous matches of the season. There is no difference in the averages between the two groups of players (-0.007, $p = 0.163$, $N = 4,628$). This result supports our assumption that conditional on hitting the post, the outcome of the shot (goal/no goal) is not informative about the player's performance.

The magnitude of this effect is large. It represents 69% of a standard deviation in individual players' ratings and 13% of a standard deviation in teams' ratings. Another way to measure this effect is to note that it represents two thirds of the average difference in ratings between players who score and player who do not score (0.710-0.531). Our result suggests that a large part of observed variations in players' ratings may be driven by the outcome bias.[12]

---

[12]Given this result, one may wonder whether journalist ratings are just noise. We show in Section 5 that these ratings are predictive of managers decisions to field players in the next match.

League		Indiv. Rating		Team Rating	
		Current Match	Year average[β]	Current Match[δ]	Year average[β]
All	Goal	0.113***	−0.007	0.015***	0.001
	p-value	< 0.001	0.163	< 0.001	0.687
	N	4,874	4,628	4,874	4,874
German Bundesliga	Goal	0.158***	−0.007	0.023*	0.005
	p-value	< 0.001	0.369	0.002	0.148
	N	2,174	2,048	2,174	2,174
French Ligue 1	Goal	0.098***	−0.002	0.011*	< 0.001
	p-value	< 0.001	0.700	0.032	0.888
	N	1,454	1,382	1,381	1,454
English Premier League	Goal	0.071***	−0.007	0.007	−0.002
	p-value	< 0.001	0.159	0.145	0.205
	N	1,246	1,198	1,246	1,246

Table 5: Effect of scoring after hitting the post on individual and team ratings. Using 1 neighbor matching and Euclidean distance. For the main estimate, the average distance between the locations of a shot and its counterfactual is 68cm. * significance at the 5% level, *** significance at the 0.1% level. [δ] excluding the player making the shot. [β] average on the previous matches of the season; 246 players hitting the post did not have ratings previously in the season.

4.3 Effect on fans' rating

On a small subsample we observe fan ratings. We find a very significant effect of similar magnitude. The effect on individual rating is 0.075 ($p < 0.001$, $N = 595$), which represents 63.1% of a standard deviation. Here again, we do not observe any effect on the previous matches of the season (0.001, $p = 0.867$, $N = 549$).[13]

It is worth noting that the bias exhibited by sport fans is of similar magnitude as the one from sport journalists. Unlike journalists who are third party observers, fans may be considered as another type of principals as their involvement in supporting their team is typically linked with their desire for the team to perform well.

5 Robustness checks

5.1 Validity of the identification strategy

Our identification strategy relies on the outcome of the shots on posts being as good as random. A casual observation of the distribution of shot trajectories in football matches suggests that players are unable to adjust shots within a precision of a few centimeters. If it was the case, one would see most shots inside the goal and right near the posts in order to be out of reach of the goal keeper. On the contrary, shots are often right at the goal keeper or way out of the frame.

In Section 4 we showed that players who scored after a shot on the post did not have higher ratings during the previous matches of the season than players not scoring

---

[13]Similarly, to Section: 4.2 we do not include the match where the opportunity happened.

after hitting the post. We provide here further evidence in support of our identification strategy. Table 6 presents a wide range of tests of balance in players's characteristics to assess whether the matching was successful. Out of all the tests, none are significant at 5%. For instance, players are not more likely to start the match instead of being a substitute when putting the ball in (-0.004, $p = 0.597$, $N = 13,066$). We also do not observe differences in the players' past performances. Since the start of the season, the players putting the ball in did not score more goals than the ones putting the ball out (0.061, $p = 0.590$, $N = 12,715$), and they did not have higher ratings on average (-0.007, $p = 0.163$, $N = 4,628$). Even more noticeably, whenever they did hit the post since the start of the season, there was no differences in propensity to put the ball in rather than out (-0.013, $p = 0.321$, $N = 5,317$). We also do not find significant differences in players' market values ($487,470$, $p = 0.203$, $N = 13,023$) and teams' market values ($311,411$, $p = 0.122$, $N = 13,066$). Nor do we find evidence that players putting the ball in were in teams more likely to win ex-ante as estimated by bookmakers betting odds (0.007, $p = 0.157$, $N = 13,063$).

	All matches			Matches with ratings		
	Diff	p	N	Diff	p	N
Player's basic characteristics						
Player starting the match	-0.004	0.597	13,066	-0.017	0.083	4,874
Forwards	0.002	0.898	13,066	0.012	0.566	4,874
Midfielder	-0.006	0.620	13,066	0.002	0.936	4,874
Defender	0.004	0.683	13,066	-0.013	0.449	4,874
Home team	-0.001	0.959	13,066	-0.014	0.490	4,874
Player's performance since the start of the season						
Number of goal scored	0.061	0.590	12,715	-0.020	0.905	4,746
Average rating	-0.007	0.163	4,628	-0.007	0.163	4,628
Number of post inside	-0.013	0.221	12,715	-0.007	0.670	4,746
Frequency of post inside	-0.013	0.321	5,317	0.011	0.616	2,049
Market values						
Player's market value	$487,470$	0.203	13,023	$-70,251$	0.867	4,860
Team's average market value	$311,411$	0.122	13,066	$-15,834$	0.952	4,874
Team's average market value $^\delta$	$302,276$	0.117	13,023	$29,015$	0.910	4,860
Opponent team's average mv	$44,617$	0.760	13,064	$-142,336$	0.504	4,874
Ex-ante probability from betting odds						
Probability to win the match	0.007	0.157	13,063	< 0.001	0.954	4,874
Probability to lose the match	-0.005	0.294	13,063	0.001	0.888	4,874

Table 6: Tests of balance of covariates between matched observations. For each variable the table presents the average difference (Diff) between matched observations, the p-value (p) of the difference from a matching estimator, and the number of observations (N). These balancing tests are done both on the whole dataset used for the study of managers' decisions and on the sample of observations used for the study of journalists ratings. One-nearest neighbor matching with Euclidean distance. $^\delta$ excluding the player hitting the post.

While there are no significant differences in the characteristics of players putting the ball in and out, some residual (non significant) differences exist. For instance, players

scoring goals are slightly more expensive than players not scoring goals. Tables 7 and 8 present the results of further analyses which control for these residual differences. In addition to matching the shots on the spatial coordinates we also match the shots on a range of other characteristics: the fact of being or not in the starting line-up; the rating the players received on the previous match; the players' average ratings since the start of the season; the players' market value; the ex-ante probability to win as predicted by the bookmaker; the players' positions. In each case, estimates stay very close from our main estimates. Our results are therefore robust to different controls of players' characteristics.

Effect of minutes played	2.68*	2.15*	2.09†	2.28*
p-value	0.013	0.043	0.053	0.031
N	12,273	12,308	12,311	12,270
Variables used for matching:				
Spatial coordinates	✓	✓	✓	✓
Starting vs. Substitute $^{\delta}$	✓	✓	✓	✓
Players' market value	✓			✓
Ex-ante probability to win		✓		✓
Position $^{\delta}$			✓	✓

Table 7: Effect of scoring on the number of minute played in the next match. $^{\delta}$ Exact matching. 1-neighbor matching with the Mahalanobis distance. † significance at the 10% level, * significance at the 5% level.

Effect of ratings	0.117***	0.121***	0.117***	0.119***	0.110***
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
N	3,845	4,628	4,860	4,874	3,832
Variables used for matching:					
Spatial coordinates	✓	✓	✓	✓	✓
Rating_{t-1}	✓				✓
Av. rating since start season		✓			✓
Players' market value			✓		✓
Ex-ante probability to win				✓	✓

Table 8: Effect of scoring on individual ratings matching on the spatial coordinates of the shots, the rating on the previous match, average rating since the start of the season, players' market value and ex-ante probability to win. $^{\delta}$ Exact matching. 1-neighbor matching with the Mahalanobis distance. *** significance at the 0.1% level.

5.2 Alternative interpretations of the results

The players' ratings are either explicitly or implicitly intended to reflect their overall performance. This is evidenced by the justifications given to such ratings when ratings are accompanied with a short text. However, journalists are not rewarded explicitly for being accurate. A possible concern about our results on journalists' ratings is that it may reflect idiosyncratic aspects of sport journalists' incentives when reporting performance and not so much a bias in their judgement. We show here that journalists' ratings are predictive of managers' future decisions to field players which suggest that they are intended to be informative of the players' performance.

We look at whether they are good predictors of the players' probability to play in the next match. Similarly, we examine the probability to be a starting player in the next match and the number of minutes played in the next match. Table 9 presents the results. We see that the journalist's rating are a good predictor of the probability to play in the next match. As managers try to select their best players, this result shows that there is a good correlation between the journalists' ratings and the managers' decisions who have well defined incentives.

	ΔPlay	ΔStart	ΔMinute
Rating	0.034***	0.045***	3.74***
p	< 0.001	< 0.001	< 0.001
N	$158,102$	$158,102$	$158,102$

Table 9: Effect of one standard deviation in journalist ratings on players' chances to be fielded in the next match. The dependent variable is the difference between the decision in match $t+1$ and in match $t-1$. The first and last match of each season are not included, standard errors clustered per player-season. *** significance at the 0.1% level

A very different concern could be that differences in ratings reflect actual differences in performance. Scoring could have an impact on the player's later performance in the match and this would be reflected in a higher rating at the end of the match. Players who scored could gain confidence and raise their game afterwards. To study this possibility we divide the match into periods of 15 minutes and test the effect for the opportunities which happens in each of these periods. The results of this analysis are displayed in Table 10. The effect is independent from when the opportunity occurs during the match. For instance, when the opportunity occurs in the additional time of the second period, the effect is 15% ($p < 0.001$, $N = 165$), for a post that occurs that late in the match the observed effect is unlikely due to a late change in performance.

	Period						
Minute	0-15	15-30	30-45[1]	45-60	60-75	75-90	90 $<$
Individual Rating	0.11***	0.14***	0.14***	0.08***	0.09***	0.10***	0.15***
p-value	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
N	668	715	922	838	825	741	165

Table 10: Effect depending on the period of the match. One-nearest neighbor matching with Euclidean distance. [1]Including the additional time of the first period. *** significance at the 0.1% level.

6   Conclusion

We have investigated a possible violation of the informativeness principle in the field: the fact that luck may be overly rewarded in performance evaluation. We isolated situations where successful or unsuccessful outcomes can be considered as good as random conditional on an observable signal of performance. We then investigated whether the outcomes of such situations influence the judgements made on the performance of the agent.

We find that the managers' decisions to field players are influenced by the shots' outcomes. Players who score in a match after hitting a post are given on average significantly more playing time in the following match. This is particularly the case when the goal may have been interpreted as being critical to the outcome of the match, moving the result from a draw to a lead at the time of the shot. On average, the effect represents 40% of the average effect of scoring a normal goal which suggests that a substantial part of the influence of players' goals on the managers' staff decision to field them in the next match could come from the outcome bias.

We find that journalists who are experts in their field are also overly influenced by the shots' outcomes. They give on average a higher rating to a player who scored after hitting the post than to a player who did not score in the same situation, for shots taken from a very close location. The effect is large as the difference in ratings represents two thirds of a standard deviation of the distribution of players' ratings. We also observe an effect of similar magnitude in ratings of sport fans

It is notable that an outcome bias is observed in such a field setting. First, agents' performances are scrutinised here to a rare degree of precision. Professional footballers' actions are followed by many cameras, recorded and commented by experts. We can expect an outcome bias to be smaller than in typical settings found in organizations where there is much less information about the agents' actions. Second, managers have high incentives to give playing time to the best performing players. Managers are typically dismissed after a period of unsatisfactory performances and the average tenure of a football manager in Europe is relatively short, 17 months (UEFA data). The pressure is high on the managers to make the best decisions. It is significant that their choice of fielding players is characterised by an outcome bias in spite of their expertise an incentives.

These elements suggest that over-rewarding luck may be widespread in the field in situations where agents actions undergo less monitoring and where decision makers' incentives are not as high to make the right decisions. The potential cost of such an outcome bias in organizations is large. It introduces inefficiencies and inequities in the allocation of sanctions, rewards, and promotions in economic institutions. For these reasons, greater care should be given in evaluation processes to limit such a bias from evaluators.

References

Abadie, Alberto, and Guido W Imbens. 2006. "Large sample properties of matching estimators for average treatment effects." *econometrica*, 74(1): 235–267.

Alicke, M.D., T.L. Davis, and M.V. Pezzo. 1994. "A posteriori adjustment of a priori decision criteria." *Social Cognition*, 12(4): 281–308.

Baron, J., and J.C. Hershey. 1988. "Outcome bias in decision evaluation." *Journal of Personality and Social Psychology*, 54(4): 569.

Bol, Jasmijn C. 2011. "The determinants and performance effects of managers' performance evaluation biases." *The Accounting Review*, 86(5): 1549–1575.

Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract theory.* MIT press.

Chapman, G.B., and A.S. Elstein. 2000. "Cognitive processes and biases in medical decision making." *Decision making in health care: Theory, psychology, and applications*, 183–210.

Coates, John. 2012. *The Hour Between Dog and Wolf: Risk Taking, Gut Feelings and the Biology of Boom and Bust.* Penguin Press.

Gauriot, Romain, and Lionel Page. 2015. "I take care of my own: a field study on how leadership handles conflict between individual and collective incentives." *The American Economic Review*, 105(5): 414–419.

Gino, F., L.L. Shu, and M.H. Bazerman. 2010. "Nameless+ harmless= blameless: When seemingly irrelevant factors influence judgment of (un) ethical behavior." *Organizational Behavior and Human Decision Processes*, 111(2): 93–101.

Hölmstrom, Bengt. 1979. "Moral hazard and observability." *The Bell journal of economics*, 74–91.

Hölmstrom, Bengt. 1982. "Moral hazard in teams." *The Bell Journal of Economics*, 324–340.

Marburger, Daniel R. 2003. "Does the assignment of property rights encourage or discourage shirking? Evidence from major league baseball." *Journal of Sports Economics*, 4(1): 19–34.

Prendergast, Canice. 1999. "The provision of incentives in firms." *Journal of economic literature*, 37(1): 7–63.

Prendergast, Canice, and Robert H Topel. 1996. "Favoritism in Organizations." *The Journal of Political Economy*, 104(5): 958–978.

Prendergast, Canice, and Robert Topel. 1993. "Discretion and bias in performance evaluation." *European Economic Review*, 37(2): 355–365.

A   Robustness check on manager decision

Table 11

	Δ Play				Δ Start				Δ Minutes			
All situations												
Effect of Scoring	0.009	0.006	0.002	0.009	0.026†	0.022	0.015	0.021	2.68*	2.15*	2.09†	2.28*
p-value	0.441	0.576	0.849	0.436	0.057	0.114	0.279	0.129	0.012	0.043	0.053	0.031
N	12,273	12,308	12,311	12,270	12,273	12,308	12,311	12,270	12,273	12,308	12,311	12,270
Loss → Loss												
Effect of Scoring	0.027	−0.019	−0.012	−0.005	0.057	−0.009	0.037	0.010	4.24	−0.80	2.41	1.73
p-value	0.607	0.704	0.804	0.918	0.244	0.849	0.475	0.826	0.271	0.830	0.537	0.612
N	935	939	939	935	935	939	939	935	935	939	939	935
Loss → Draw												
Effect of Scoring	−0.028	0.007	−0.032	−0.004	0.010	0.036	0.011	0.045	2.46	4.09	1.46	4.70†
p-value	0.324	0.792	0.287	0.889	0.781	0.285	0.748	0.177	0.362	0.104	0.578	0.071
N	2,141	2,148	2,148	2,141	2,141	2,148	2,148	2,141	2,141	2,148	2,148	2,141
Draw → Win												
Effect of Scoring	0.029†	0.043*	0.026	0.026	0.060**	0.074***	0.060**	0.059**	4.49**	5.34**	4.54**	4.30**
p-value	0.090	0.011	0.158	0.132	0.004	<0.001	0.006	0.004	0.006	0.001	0.008	0.008
N	5,624	5,639	5,641	5,622	5,624	5,639	5,641	5,622	5,624	5,639	5,641	5,622
Win → Win												
Effect of Scoring	0.010	0.017	0.008	0.019	−0.010	−0.001	−0.023	0.004	0.51	1.59	−0.01	1.26
p-value	0.613	0.372	0.687	0.336	0.692	0.953	0.342	0.866	0.786	0.380	0.995	0.484
N	3,573	3,582	3,583	3,572	3,573	3,582	3,583	3,572	3,573	3,582	3,583	3,572
Variables used for matching:												
Spatial coordinates	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Starting vs. Substitute δ		✓	✓	✓		✓	✓	✓		✓	✓	✓
Players' market value			✓	✓			✓	✓			✓	✓
Ex-ante probability to win				✓				✓				✓
Position δ				✓				✓				✓

Table 11: Effect of scoring on the number of minutes played in the next match. δ Exact matching. 1-neighbor matching with the Mahalanobis distance. † significance at the 10% level, * significance at the 5% level, ** significance at the 1% level, *** significance at the 0.1% level

B   Effect on journalists ratings as a function of the match outcome

	Loss → Loss	Loss → Draw	Draw → Win	Win → Win
	Individual Rating			
Effect of Scoring	0.129***	0.137***	0.125***	0.101***
p	< 0.001	< 0.001	< 0.001	< 0.001
	Team Rating			
Effect of Scoring	−0.002	0.045***	0.016**	0.015*
p	0.921	< 0.001	0.002	0.014
N	363	842	2,317	1,352

Table 12: Effect of scoring after hitting the post on individual and team ratings depending on the effect of the post on the match outcome, at the time of the shot. Using 1 neighbor matching and Euclidean distance. * significance at the 5% level, ** significance at the 1% level , *** significance at the 0.1% level. $^{\delta}$ excluding the player making the shot.