University of Zurich[UZH]

Department of Business Administration

UZH Business Working Paper Series

Working Paper No. 402

Does Performance Pressure Accentuate Outcome Bias? Evidence from Managerial Dismissals

Fabienne Jedelhauser, Raphael Flepp, Pascal Flurin Meier, and Egon Franck

May 2024

University of Zurich, Plattenstrasse 14, CH-8032 Zurich,
http://www.business.uzh.ch/forschung/wps.html

UZH Business Working Paper Series

Contact Details

Fabienne Jedelhauser

fabienne.jedelhauser@business.uzh.ch

Raphael Flepp

raphael.flepp@business.uzh.ch

Pascal Flurin Meier

pascal-flurin.meier@business.uzh.ch

Egon Franck

egon.franck@business.uzh.ch

University of Zurich

Department of Business Administration

Plattenstrasse 14, CH-8032 Zurich, Switzerland

Does Performance Pressure Accentuate Outcome Bias? Evidence from Managerial Dismissals

Fabienne Jedelhauser, Raphael Flepp, Pascal Flurin Meier, and Egon Franck*

May 2024

Abstract

Outcome bias refers to the tendency to overweight the informativeness of observed outcomes in evaluations, consequently underestimating the influence of luck. However, observed outcomes that fall short of expectations simultaneously trigger performance pressure, potentially reinforcing outcome bias in evaluation decisions such as managerial dismissals. Using data from European football, we investigate whether managerial dismissal decisions are influenced by luck operationalized as opponent player injuries and whether this influence is more pronounced under performance pressure. Our findings reveal that luck significantly impacts dismissal decisions, particularly as performance pressure mounts. Importantly, this amplified outcome bias under performance pressure is predominantly driven by instances of bad luck. These results suggest that the extent of outcome bias has been underappreciated, especially in situations involving bad luck.

Keywords
Outcome Bias, Luck, Performance Pressure, Managerial Dismissal, Principal-Agent Setting

JEL Classification
D81, D86, D91, J44, Z2

* Corresponding author: Fabienne Jedelhauser, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland, fabienne.jedelhauser@business.uzh.ch/Coauthors: Raphael Flepp, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland; Pascal Flurin Meier, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland; Egon Franck, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

1 INTRODUCTION

Performance evaluation in a principal-agent setting is a complex task since actual performance is often ambiguous and difficult to observe. Agents (e.g., CEOs) should not be punished or rewarded for factors beyond their control (Holmström, 1979), such as luck. However, in reality, "rewards for good luck" and "punishments for bad luck" may nevertheless occur. This is because observed outcomes are usually a function of skill and luck (Gauriot & Page, 2019; Rubin & Sheremeta, 2016), and principals (e.g., the board of directors) tend to underestimate the role of luck factors and overweight the observed outcome (Allison et al., 1996). Thus, principals succumb to an "outcome bias" by rewarding (punishing) agents for good (bad) luck (e.g., Baron & Hershey, 1988; Brownback & Kuhn, 2019; Gauriot & Page, 2019).

This tendency to overweight the observed outcome and thus fail to filter out luck in evaluations is widely documented in both experimental studies (e.g., Brownback & Kuhn, 2019; Gurdal et al., 2013; König-Kersting et al., 2021) and in the field (e.g., Gauriot & Page, 2019; Kausel et al., 2019). However, an often overlooked consideration in the prior literature is that observed outcomes may simultaneously prompt the emergence of performance pressure when current performance deviates from organizational expectations. The role of performance pressure seems crucial since high-stakes principal-agent evaluations, such as managerial dismissals, often occur in situations in which organizational performance is poor (Adams et al., 2010; Hilger et al., 2013; Jenter & Kanaan, 2015) or below organizational expectations (Farrell & Whidbee, 2003; Puffer & Weintrop, 1991). Moreover, performance pressure is a unique source of stress (Mitchell et al., 2019) and is prevalent across many workplace contexts (Dohmen, 2008), making it ubiquitous in organizations (Spoelma, 2022). The effect of outcome-induced performance pressure on potentially outcome-biased evaluations has yet to be explored. Thus, in this paper, we pose the following question: How does performance pressure affect outcome bias in principal-agent evaluation decisions? Therefore, we aim to obtain a more comprehensive and accurate understanding of the true effect of outcome bias.

Grounded in the idea that principals may make more accurate evaluation decisions under low pressure but more biased decisions under high pressure (Byrne et al., 2015), our core conjecture is that principals' tendency to succumb to outcome bias is more pronounced when outcome-induced performance pressure mounts since principals are distracted and preoccupied with other thoughts (e.g., Lewis & Linder, 1997). Under these circumstances, principals' ability to filter out luck is impaired, and the salience of observed outcomes (even if confounded by luck) increases, potentially reinforcing the overweighting of observed outcomes.

We focus on the evaluation decision of managerial dismissals, which is one of the most important decisions made by boards acting as principals. The decision to dismiss the manager is a direct reflection of the board's evaluation of the manager: If the board's estimated ability of the manager falls below some threshold, the manager will be dismissed (e.g., Jenter & Kanaan, 2015). Thus, we propose that luck decreases the probability of managerial dismissal and that performance pressure accentuates the consideration of luck in this relationship. Moreover, since luck can manifest as either good or bad (Liu & de Rond, 2016), we argue that both the tendency to penalize (dismiss) agents for bad luck and to reward (not dismiss) them for good luck is accentuated when performance pressure mounts.

We test our predictions in the context of European football, which allows us to overcome various challenges associated with empirical examinations of how economic actors behave in evaluation decisions (Gauriot & Page, 2019; Lefgren et al., 2015). Professional sports are a high-stakes environment and an ideal setting for testing hypotheses about sophisticated economic actors (e.g., Kahn, 2000). First, managerial dismissals in European football are high-stakes decisions. Similar to CEO dismissals that expose principals (i.e., the board of directors) to external and internal scrutiny (Hilger et al., 2013), head coach dismissals receive considerable domestic and worldwide media attention, bear substantial financial consequences, and trigger the risk of reputational damage for both the club boards and the coach (van Ours & van Tuijl, 2016). Second, weekly available performance signals in European football as opposed to less frequently available quarterly or annual performance data in a managerial context, allow for a more precise alignment of skill and luck components to turnover decisions. Third, since European football is a low-scoring sport, match outcomes are heavily influenced by luck and may not reliably reflect team performance on the pitch (Brechot & Flepp, 2020; Wunderlich et al., 2021). Consequently, observed outcomes encompass a significant luck component, which is central to our arguments.

Importantly, our setting allows us to overcome the typical measurement difficulties for luck by employing key opponent injuries as a novel source of luck. Notably, while injury-related absences of key focal players may not be completely uninformative of the coach's ability, injury-related absences of key opponent players represent a source of exogenous performance boosts for the focal team and coach. We measure luck by accumulating and median-adjusting key opponent injuries and distinguish bad luck (good luck) by determining the extent to which luck is negative (positive).[1] Finally, the context of European football allows simple distinctions between low- and high-pressure situations. Our setting permits us to proxy team-specific variations in performance pressure by calculating the difference between the league table rank and preseason targets.

Consistent with our predictions, the empirical findings reveal outcome-biased dismissal decisions: Lucky head coaches are less likely to be dismissed. Importantly, club boards are more likely to fall prey to outcome bias in situations in which observed outcomes trigger performance pressure. Specifically, outcome-induced performance pressure reduces the probability of head coach dismissals when teams experience luck. These effects, however, are primarily driven by bad luck. Thus, head coaches are particularly penalized for bad luck, and the bad luck effect is accentuated under greater performance pressure. In contrast, we do not find evidence for a reinforcing effect of performance pressure on the consideration of good luck in head coach dismissals. Taken together, these findings provide support for our conjecture that performance pressure is a key contingency factor for outcome bias.

Our findings first and foremost contribute to the literature on outcome bias by emphasizing and providing statistical evidence that ignoring observed outcomes' impact on outcome-induced performance pressure leads to an underappreciation of outcome bias. Second, by showing that luck influences managerial dismissal decisions, we add to the broader literature on the influence of luck and its effect on decision makers (e.g., Denrell et al., 2019; Liu & de Rond, 2016). More specifically, we also contribute to the literature on coach dismissals in European football by highlighting that coaches

[1] Doing this represents an additional advantage of our setting. While previous field studies on outcome-biased decisions predominantly focused on quasiexperimental regression discontinuity designs to capture luck (Flepp & Meier, 2024; Gauriot & Page, 2019; Kausel et al., 2019; Lefgren et al., 2015; Meier et al., 2022; Meier et al., 2023), we are able to quantify the extent of luck.

are dismissed for factors beyond their control. Finally, we make a methodological contribution by developing a novel measurement for luck in European football.

The remainder of this paper is structured as follows. Section 2 presents an overview of the literature and derives testable hypotheses. Section 3 describes the data and explains the estimation strategy, whereas Section 4 depicts the results of our empirical analyses and employs alternative specifications. We discuss the implications of our results and conclude the paper in Section 5.

2 RELATED LITERATURE AND HYPOTHESES

2.1 Outcome bias in principal-agent evaluations

Outcome bias is a widespread, empirically examined phenomenon that occurs when individuals overweight the informativeness of outcomes (e.g., success or failure) in evaluations (Baron & Hershey, 1988). Since taking into account observed outcomes is not generally considered outcome bias, cognitive distortion occurs only when inferences about ability are based on outcomes influenced by exogenous factors outside an individual’s control (Baron & Hershey, 1988). In such situations, the outcome is partly uninformative and obscures valuable information about an agent. The seminal work of Baron and Hershey (1988) investigates outcome bias among students evaluating medical procedures. Despite knowing the predetermined success probability of a medical procedure, students assessed the decisions more favorably after successful than unsuccessful procedures.

Falling prey to outcome bias thus describes the fallacy that good outcomes follow from good decisions and bad outcomes result from bad decisions. However, outcomes usually consist of a skill component and exogeneous factors (i.e., luck) (Gauriot & Page, 2019; Rubin & Sheremeta, 2016), and agents’ successes or failures are driven by both actions and random circumstances (Gauriot & Page, 2019). Since disentangling luck and skill is challenging (Budescu & Bruderman, 1995), individuals tend to underestimate external factors. Thus, in principal-agent evaluations, outcome-biased evaluations may result from a misattribution of luck for skill, fostering the biased belief that lucky agents are more skilled than unlucky ones (Brownback & Kuhn, 2019).

The existence of outcome bias has been documented in numerous experimental and field studies, including principal-agent settings.[2] In the laboratory experiment of Marshall and Mowen (1993), principals tend to rate the competence of an agent more positively (negatively) when outcomes are positive (negative) regardless of whether luck influences the outcome. In a role-play experiment with students as sales managers, the sales manager considered only the outcome when evaluating salespeople, independent of the appropriateness of the decision by salespeople (i.e., choosing the recommended vs. not recommended option by the sales department). Gurdal et al. (2013) find that when agents choose between safe choices or a lottery, agents are penalized based on the outcome of the lottery although being an event beyond their control. Moreover, outcome bias even occurs in transparent settings, and agents are punished for bad luck, although both effort and luck are perfectly observable (Brownback & Kuhn, 2019). In the context of predefined probabilities of success in risky and nonrisky financial investment decisions, principals reward positive outcomes even if they initially did not agree with the agent's investment strategy (König-Kersting et al., 2021).

Field evidence of outcome bias in principal-agent settings stems from sports and the managerial context. Examining the quasiarbitrary outcome of whether shots that hit a post result in a goal in European football, Gauriot and Page (2019) find that coaches overly reward good luck in performance outcomes. In the match following lucky successes, players receive more playing time and have a higher probability of being fielded and selected for the starting lineup.[3] Focusing on CEO dismissals, Jenter and Kanaan (2015) find that CEOs are dismissed after poor firm performance caused by factors beyond their control. Similarly, Flepp and Meier (2024) show that CEOs of firms that barely outperform the S&P 500 index are less likely to get fired than are CEOs of firms that barely underperform the S&P 500 index.

[2] The empirical literature also has reported evidence of outcome bias in contexts other than principal-agent settings, such as for example, self-evaluations (Lefgren et al., 2015; Meier et al., 2022; Meier et al., 2023), independent third-party evaluations (Brownback & Kuhn, 2019; Kausel et al., 2019), or betting markets (Flepp et al., 2023).

[3] Lefgren et al. (2019) investigate circumstances in the NBA in which a discrete performance outcome (winning/losing), although being partly uninformative about an agent's skills, contains relevant information about an agent and is arguably in line with optimal contracting. However, part of their results could also be attributed to outcome bias.

Applying the above-documented arguments and findings to the high-stakes evaluation decision of managerial dismissal suggests that principals may erroneously perceive agents encountering higher levels of luck as more skilled than what the true underlying performance actually suggests. Consequently, agents are likely to be "rewarded" with a lower dismissal probability when they experience more luck. Since luck has a dual nature by being either good or bad (Liu & de Rond, 2016), good luck captures beneficial events leading to desirable outcomes, whereas bad luck is associated with unpleasant outcomes (Baumeister et al., 2001). Positive events beyond an agent's and an organization's control may thus be considered good luck, whereas negative events imply bad luck (Amore & Schwenen, 2022). Since bad luck is more likely to lead to worse performance outcomes than is good luck, we assume that principals erroneously perceive agents experiencing bad luck as less skilled and those experiencing good luck as more capable. Overall, we argue that agents are penalized for bad luck and rewarded for good luck. We thus state the following hypotheses:

Hypothesis 1a: *A higher level of luck is associated with a decreased probability of managerial dismissal.*

Hypothesis 1b: *A higher level of bad luck is associated with an increased probability of managerial dismissal.*

Hypothesis 1c: *A higher level of good luck is associated with a decreased probability of managerial dismissal.*

2.2 Moderating role of performance pressure

The extant literature on outcome bias typically has considered how observable outcomes influence principals' evaluation of an agent. Therefore, the focus has been only on one side of the coin. However, the neglected second side concerns the influence of observable outcomes on performance pressure, which has been largely overlooked. Since organizations are driven by different expectations (meeting sales targets, winning the league title, etc.), observed outcomes relative to organizational expectations may play a crucial role in determining whether current performance is deemed satisfactory and, thus, whether principals face performance pressure. In particular, high-stakes evaluation decisions such as managerial dismissals are often associated with unsatisfactory performance relative to organizational expectations (Farrell & Whidbee, 2003; Puffer & Weintrop, 1991) and, thus, are (arguably) typically made under performance pressure. The effect of such outcome-induced performance pressure on potentially outcome-biased evaluations has yet to be explored. In the following, we outline why performance pressure is an important contingency for outcome bias.

We start by asserting that performing below organizational expectations leads to the emergence of performance pressure. Performance pressure generally describes a situation in which the importance of performing well is heightened (Baumeister, 1984). Building upon this concept, Gardner (2012) further characterizes performance pressure as a situation under heightened scrutiny that demands high performance due to significant consequences of failing to meet expectations, such as, for example, the threat of job termination if clients of consulting or audit teams are not satisfied.[4] Importantly, the urgency to improve performance due to potential negative consequences heavily affects individuals and creates pressure (Mitchell et al., 2019). The experience of such performance pressure then demands responses to tackle potential threats stemming from, for example, failing to meet expected performance standards (Spoelma, 2022). Hence, a situation in which observed outcomes fall short of organizational expectations captures a performance pressure scenario in line with common understandings of performance pressure.

Although performance pressure has also been associated with increased task performance through motivational factors (Eisenberger & Aselage, 2009; Gardner, 2012), understanding its effect on potentially biased decision making in a high-stakes environment requires focusing on the influence that performance pressure has on cognitive capacities and information processing (Byrne et al., 2015; Lewis & Linder, 1997). Distraction theory with respect to cognitive abilities (Byrne et al., 2015; Lewis

[4] Gardner (2012) further identifies shared accountability as an important interrelated factor creating performance pressure on teams. Principals in our settings are club boards, which are likely to be comprised of multiple individuals. However, the underlying reasoning remains similar even if club boards were composed of a single individual.

& Linder, 1997; Wine, 1971) suggests that performance pressure impairs task performance by increasing cognitive load due to increased distraction and thus preoccupation with other thoughts.[5]

Specifically, performance pressure may decrease task performance by redirecting attentional focus away from task-relevant information. Distraction, however, can arise from both external and internal factors. For instance, internal anxiety potentially resulting from performance pressure drives attention away from primary tasks and impairs performance since individuals are preoccupied with other thoughts (e.g., reducing their anxiety) (Lewis & Linder, 1997). Alternatively, distraction may arise from external factors (Byrne et al., 2015; Lewis & Linder, 1997). For example, stakeholders (e.g., fans and media)[6] who demand corrective actions to address unsatisfactory performance shortfalls could distract the information processing of principals. Addressing stakeholders' inquiries (e.g., press conferences or stakeholder meetings) may similarly foster principals' preoccupation with other concerns, increasing the salience of observed outcomes. Thus, both internal and external distractions likely appear in tandem or reinforce each other, potentially impairing the cognitive ability to process task-relevant information.

Taken together, our core arguments thus build on the premise that performance pressure creates mental distractions that may impair cognitive abilities (Byrne et al., 2015). Indeed, acute stressful situations can restrict the information processing of individuals and groups by narrowing attention and simplifying information (Staw et al., 1981), and the cognitive ability to perform complex tasks diminishes under scrutiny (Ellis, 2006). Similarly, time constraints and uncertainty impair the capacity to systematically and deliberately process information (Kahneman, 2003, 2011; Soll et al., 2015), and the demand for quick actions redirects the cognitive system to find fast solutions (e.g., Dreu, 2003;

[5] Vast literature has explained the underlying reasons individuals' task performance is reduced under pressure (Baumeister, 1984; Baumeister & Showers, 1986; Böheim et al., 2019; Byrne et al., 2015), predominantly referred to as "choking under pressure" in motor tasks (Baumeister, 1984; Böheim et al., 2019; Dohmen, 2008; Gray, 2004; Harb-Wu & Krumer, 2019). However, since we focus on the impact of performance pressure on the (cognitive) decision-making capacity of principals, we do not refer directly to and thus differ from the literature on "choking under pressure", although distraction theory can also explain why individual performance is suboptimal under pressure in motor tasks. In contrast, we relate our paper rather to the literature examining "choking" in cognitive tasks, such as, for example, performance declines in math tests due to test anxiety (Wine, 1971).

[6] In an organizational context, pressure to find solutions when performance falls short of expectations could stem from (institutional) shareholders (Fisman et al., 2014).

Finucane et al., 2000; Gilbert & Hixon, 1991). Thus, given performance pressure, shifts in focus from task-relevant to the consideration of task-irrelevant information are likely (Lewis & Linder, 1997).

The increased tendency to consider irrelevant information under performance pressure has important implications for outcome-biased evaluations. When individuals' cognitive abilities are not encumbered by situational demands, outcome-biased inferences are less likely (Allison et al., 1996). As performance pressure increases, principals tend to overweight observed outcomes more in high-pressure situations than they would in low-pressure situations, suggesting that they may unconsciously assign greater importance to luck in observed outcomes. Thus, by reducing cognitive resources, performance pressure may increase principals' tendency to fall prey to outcome bias in evaluations. As a result, we argue that principals facing greater performance pressure are, due to cognitive constraints, less capable of filtering out luck in performance outcomes, leading to an increased probability of considering luck in managerial dismissal decisions.

Regarding the effect of performance pressure on the consideration of bad and good luck, we argue that bad luck potentially increases outcome-induced performance pressure. This, in turn, may result in an overemphasis on bad luck in observed outcomes and, consequently, an accentuated probability of managerial dismissal in instances of bad luck and performance pressure. Conversely, if outcomes, though below expectations, involve good luck, the probability of too mild performance pressure heightens. Since performance pressure nonetheless leads to preoccupation with other thoughts, principals may become too lenient in their evaluations by shifting the focus to the observed outcomes, and are more likely to consider good luck. Consequently, performance pressure amplifies the negative effect of good luck on the probability of managerial dismissal.

Overall, we conjecture the following in the high-stakes evaluation decision of managerial dismissal:

Hypothesis 2a: *The negative relationship between luck and the probability of managerial dismissal is accentuated under higher levels of performance pressure.*

Hypothesis 2b: *The positive relationship between bad luck and the probability of managerial dismissal is accentuated under higher levels of performance pressure.*

Hypothesis 2c: *The negative relationship between good luck and the probability of managerial dismissal is accentuated under higher levels of performance pressure.*

3 METHODS

3.1 Data and sample

We test our hypotheses using a comprehensive dataset on the top five European football leagues (i.e., Premier League, La Liga, Serie A, Ligue 1, and 1. Bundesliga) between the 2016/2017 and 2020/2021 seasons. We collect various information on players, coaches, and teams from the website www.transfermarkt.com. In particular, we derive team compositions, players' injury histories, starting lineups, corresponding coaches, and historical end-of-season league table standings. We further rely on the website www.kicker.de to obtain match week based league table ranks. The primary dataset covers 18,058 team performance observations from 9,029 games.[7] As we discuss below, we identify 149 head coach dismissals and exclude all team-season observations after the first dismissal, resulting in 15,458 team performance observations for our main analysis.

3.2 Variables

3.2.1 Dependent variable

In line with Pieper et al. (2014) and Flepp and Franck (2021), we define within-season coach changes as dismissals if the associated (online) press articles suggest that the turnover was involuntary. We ignore dismissals in the first four games since multiple teams often have the same league table standing, and a meaningful assessment of a team's performance relative to preseason expectations is difficult.

Further building on Besters et al. (2016) and Flepp and Franck (2021), we consider only the first within-season coach dismissal.[8] The underlying reasoning is twofold. On the one hand, we want to ignore the dismissal of a caretaker who might be replaced after a few matches (Besters et al., 2016). On

[7] Every game appears twice in our analysis, once from the home team perspective and once from the away team perspective.
[8] Our results are robust to the consideration of all dismissals (full sample).

the other hand, we want to ensure that our (cumulative) luck measurement is only attributable to the specific coach. Overall, we registered 149 coach dismissals.

3.2.2 Luck

By definition, luck must be independent of the coach's quality and external to the coach and team (Gauriot & Page, 2019; Rubin & Sheremeta, 2016). Building on the finding that injury-related absences of key players decrease the focal team's performance in general (Jedelhauser et al., 2023), we argue that injury-related absences of key opponent players are a source of exogeneous performance boosts that increase the winning chances of the focal team. While injuries of the focal team may not be completely uninformative of the coach's ability (e.g., due to intense training), opponent injuries are arguably exogeneous to the focal team and coach. Therefore, we assume that a team will benefit from luck throughout the season when it gains more advantages from opponent injuries than other teams.

To build our luck measurement, we first identify key opponent players by relying on the number of games players have been selected for the starting lineup within a season and team. Coaches aim to start with their strongest players at each position to increase the probability of winning the game, and selection to starting lineups captures the relative importance of players (Gauriot & Page, 2019; Lefgren et al., 2015). Thus, key players are defined as those who have been selected to the starting lineup in more than 90%[9] of possible games per team and season.[10] We identify 829 key players out of 4,636 players who have at least once been selected for the starting lineup. Of the 829 identified key opponent players, 551 reported at least one injury-related absence throughout our observation period.

Second, we proceed by aggregating the number of key opponent injuries to the focal team's game level. In particular, we sum the number of key opponent injuries the focal team profited from until and including the focal game in a particular season. We derive the cumulative number of key opponent injuries as a source of exogenous performance boosts throughout the season. For example, if Arsenal's

[9] Our results do not change when we apply an alternative threshold of 80%.

[10] The number of possible games per player and team is adapted in two ways. First, since we are aware of 401 within season player changes in the top five leagues, a player can be considered a key player for two teams within one season if, for example, he changed teams during the winter transfer window. We manually check these players' transfer histories to adjust the number of possible games per player, team, and season. Second, the number of possible games is injury-adjusted. We reduce the number of possible games by the number of games a player missed per team due to an injury-related absence.

opponent in round one had two key players injured and the opponent in round two had three key players injured, the cumulative number of key opponent injuries have a value of five after Arsenal's second game.

Third, every team is naturally expected to profit from key opponent injuries to a certain extent. Thus, we subtract the median value of the cumulative key opponent injuries per league, season, and games played from the focal team's cumulative number of key opponent injuries to construct our *Luck* measurement.[11] Doing so allows us to distinguish team-specific levels of luck relative to the expected values of all teams within a league and season. The advantage of this approach is that the *Luck* variable takes on positive or negative values and does not systematically increase over time.

3.2.3 Bad and good luck

To distinguish bad and good luck, we split *Luck* into two components. Positive values of *Luck* suggest that teams profited more from key opponent injuries than did the median team. Thus, teams with positive values were relatively more lucky. In contrast, negative values of *Luck* indicate that teams profited less than the general expectation; thus, the teams experienced relatively bad luck. Specifically, we distinguish teams with relatively bad luck from those with relatively good luck by constructing two semicontinuous variables as formally expressed in equation (1).[12] *Bad Luck* equals the absolute value of *Luck* if *Luck* is negative; otherwise, it is equal to zero. *Good Luck* equals *Luck* if *Luck* is greater than or equal to zero; otherwise, it is equal to zero:

$$\begin{aligned} &Bad\ Luck: |Luck|\ if\ Luck < 0, zero\ otherwise \\ &Good\ Luck: Luck\ if\ Luck\ \geq 0, zero\ otherwise \end{aligned} \tag{1}$$

According to our conceptualization of luck, the measure is valid if it is external to the focal coach and club and if it impacts observed outcomes in European football. For the former, coaches likely have no influence over the opponent team's prior injuries. For the latter, Table A2 validates that our conceptualization of luck captures exogenous performance boosts. Columns (1) and (2) provide empirical evidence that the number of *Key Opponent Injuries* significantly increases team performance

[11] The interpretation of our results remains similar if we take the mean instead. However, we opted for the median as it is less sensitive to outliers than is the mean.

[12] Table A1 presents a real data example to illustrate the construction of our main independent variables.

in a single game as proxied by the *Goal Difference* (column 1) and *Points* won (column 2) in a single game.[13] Crucially, column (3) reveals that the accumulated and median-adjusted *Luck* variable significantly reduces the league table rank. Thus, experiencing relatively more luck leads to better performance. Finally, *Bad Luck* increases the league table rank, whereas *Good Luck* is associated with lower ranks in column (4).[14]

3.2.4 Performance pressure

One advantage of our settings is the clear distinction between high- and low-pressure situations. We operationalize performance pressure to be present when performance falls short of organizational expectations. Since organizations often build their target based on historical performance, we construct a team's preseason expectations based on the weighted exponential moving average of its last season rank and its previous historical aspirations (*HAL*) (e.g., Moliterno et al., 2014), as indicated in equation (2):

$$HAL_t = \alpha \times Final\ Rank_{t-1} + (1-\alpha) \times HAL_{t-1} \qquad (2)$$

The intuition behind *alpha* is that high alphas indicate greater importance of more recent performance outcomes, whereas lower alphas increase the importance of more distant outcomes. We set alpha equal to one in our main analysis because European football is characterized by a short-term memory of historical performance (Moliterno et al., 2014)[15]. Thus, *HAL* corresponds to a team's final rank in the previous season.[16] To construct the *Performance Pressure* variable, we calculate the rank difference between a team's rank (*Rank*) after each match and a team's historical aspirations (*HAL*), as formally expressed in equation (3). Positive values of *Performance Pressure* indicate that teams are below preseason expectations and thus the extent of performance pressure. Negative values are an indication of the extent to which a team currently overperforms:

$$Performance\ Pressure = Rank - HAL \qquad (3)$$

[13] These proxies have previously been used to measure team performance in a single game in European football (e.g., Flepp & Franck, 2021; Jedelhauser et al., 2023; T. Peeters, 2018)
[14] Note that in European football, high ranks refer to bad performance and low ranks to good performance.
[15] In a robustness test (outlined in Section 4.2), we test for an alternative alpha of 0.75, which puts a higher weight on past performance outcomes.
[16] Teams that got promoted received the highest possible rank within the league as the final rank in the previous season.

3.2.5 Controls

Although our *Luck* variable is exogenous and hence not assumed to be correlated with team characteristics, we include several controls. First, as performance outcomes matter for dismissal decisions in general (Bryson et al., 2021; Pieper et al., 2014; van Ours & van Tuijl, 2016), we control for the cumulative percentage of points a team has achieved out of possible points, including the focal game (*Point Percentage*)[17]. Second, we control for *Tenure* in days at the current club, as the experience of a coach could protect him from dismissal. Moreover, longer tenure at a club may facilitate disentangling luck and skill, as club boards become more acquainted with the coach. Third, we control for *Games Played*, as coach dismissals may not be equally distributed throughout the season (Pieper et al., 2014). Indeed, in our sample, 67% of the coach dismissals occurred in the first half of the season.

Fourth, although key opponent injuries are exogenous, incentives to strategically rest key players before important games could affect key opponent injuries and thus our luck measurement. Important games, such as UEFA competitions, can create greater incentives to strategically rest key players (Jedelhauser et al., 2023; Kaplan, 2022; T. Peeters, 2018). We thus include a dummy variable that captures whether the opponent is still competing in UEFA competitions (*UEFA Period Opponent*). Fifth, as performance outcomes might be affected by the focal team's own injury-related absences (Jedelhauser et al., 2023), we control for the median-adjusted cumulative number of key focal player injuries (*Key Focal Injuries*).[18]

Finally, to account for unobserved but time-constant within-season differences in team quality, we add team-season fixed effects (Flepp & Franck, 2021; van Ours & van Tuijl, 2016). We also include coach fixed effects to control for unobserved, stable coach characteristics. This is crucial for alleviating the potential concern that, although good and bad luck can happen to anyone, certain individuals may have superior strategic responses to luck (de Rond, 2014; Liu & de Rond, 2016).

Table 1 depicts the summary statistics and correlations of the main variables of interest. The creation of the *Luck* variable is based on the full data sample (18,058 observations); however,

[17] A common option in other sports than football is to rely on cumulative win percentage (Lefgren et al., 2019). However, since the rules in European football accept and reward draws with one point, the cumulative percentage of points achieved out of all possible points seems most appropriate.
[18] The median is again taken per league, season, and games played.

descriptive statistics are calculated for the observations included in our main analyses (15,458 observations).

Insert Table 1 about here

3.3 Model specification

We test our hypotheses by running OLS regressions. An OLS estimation to explain a binary outcome is an appropriate choice if the interest is on marginal effects (Wooldridge, 2010; Wulff et al., 2023). Moreover, the interaction terms in logit models are more difficult to interpret than are those in OLS models (Hoetker, 2007). In our baseline OLS regression analysis, we thus regress *Coach Dismissal* on *Luck*. We then add our set of controls and fixed effects, as formally expressed in equation (4), where *i* denotes the game, *j* the team, *k* the season, and *l* the coach. *X* is a vector of control variables, η contains team-season fixed effects, and θ captures coach fixed effects. We adjust for potential serial correlation in panel data by computing standard errors clustered at the game level:[19]

$$Coach\ Dismissal_{ijkl} = \beta_0 + \beta_1 \times Luck_{ijkl} + \beta X + \eta_{jk} + \theta_l + \varepsilon_{ijkl} \quad (4)$$

To identify whether the consideration of luck is accentuated under performance pressure, we interact *Performance Pressure* with *Luck* in equation (5).

$$Coach\ Dismissal_{ijkl} = \beta_0 + \beta_1 \times Luck_{ijkl} + \beta_2 \times Performance\ Pressure_{ijkl} + \beta_3 \times Performance\ Pressure_{ijkl} \cdot Luck_{ijkl} + \beta X + \eta_{jk} + \theta_l + \varepsilon_{ijkl} \quad (5)$$

4 RESULTS

4.1 Main results

Table 2 reports the results of testing our hypotheses. In columns (1) and (2), we test Hypothesis 1a. In column (1), we include only *Luck* as an explanatory variable. The *Luck* coefficient is negative and significant (β = -0.00141, p value < 0.01), implying that coaches of teams with relatively higher levels of luck have a lower probability of being dismissed. We include team-season fixed effects, coach

[19] Clustering the standard errors at the team level does not change our results.

fixed effects, and our set of control variables in column (2), and the coefficient of *Luck* remains negative and significant (β = -0.00176, p value < 0.05). Thus, columns (1) and (2) support Hypothesis 1a, which states that higher levels of luck decrease the probability of managerial dismissal.

Importantly, this effect is also economically relevant. A one standard deviation increase in *Luck* (1.556) decreases the probability of dismissal per matchday by 0.27 percentage points, which corresponds to a decrease in the probability of dismissal by 27% compared to the average dismissal rate of 0.01 per matchday (column 2).

Columns (3) and (4) reflect the test of Hypotheses 1b and 1c. The baseline estimation in column (3) reports a negative and statistically significant effect of *Bad Luck* on the probability of managerial dismissal (β = 0.00273, p value < 0.05). We include fixed effects and controls in column (4). The *Bad Luck* coefficient has a positive and significant effect (β = 0.00434, p value < 0.05) on the probability of managerial dismissal, whereas *Good Luck* remains insignificant (β = 0.0002, p value > 0.98). Since disentangling the effects of bad and good luck suggests that the effects of luck on the dismissal probability are mainly driven by bad luck, we find evidence for Hypothesis 1b but lack statistical support for Hypothesis 1c. Regarding the economic relevance of the *Bad Luck* coefficient in column (4), the regression analysis indicates that a one standard deviation increase in *Bad Luck* (0.826) corresponds to an increased probability of dismissal of approximately 36% per match day compared to the average dismissal rate.

Insert Table 2 about here

Column (5) reflects the test of Hypothesis 2a, which states that performance pressure accentuates the consideration of luck in evaluations. Column (5) shows that the interaction coefficient of *Performance Pressure* and *Luck* is negative and statistically significant (β = -0.00055, p value < 0.01). Thus, and in line with Hypothesis 2a, we find statistical support that principals under performance pressure tend to be more outcome-biased in their agent evaluations.

Column (6) of Table 2 reflects our test of Hypotheses 2b and 2c. Column (6) shows that the effect of *Bad Luck* is accentuated when performance pressure increases, as indicated by the positive and significant interaction coefficient between *Performance Pressure x Bad Luck* (β = 0.00150, p value < 0.001). In contrast, our data do not provide statistical support for the reinforcing effect of performance pressure on the consideration of good luck in dismissal decisions. Column (6) shows that the interaction effect of *Performance Pressure x Good Luck* (β = 0.00016, p value > 0.36) is positive but remains insignificant. Thus, although we find supporting statistical evidence for Hypothesis 2b, we fail to find support for Hypothesis 2c. The results reveal that the reinforcing effect of performance pressure on outcome bias in head coach dismissal decisions is particularly pronounced on club boards' responses to bad luck. Consequently, performance pressure has a reinforcing effect solely in instances of bad luck but not when coaches experience good luck.

4.2 Robustness tests

In our main analysis, we implicitly assume that coaches receive a new start every season. To soften this assumption, we estimate coach-specific duration models in which the tenure of a coach relates to separate coach-team dyads (van Ours & van Tuijl, 2016).[20] This further allows us to compare our results to frequently used duration models in the sports context (e.g., Bryson et al., 2021; Semmelroth, 2022; van Ours & van Tuijl, 2016). Table 3 reports the results of our right-censored duration analysis based on Cox proportional hazard models.

The estimations support our previous findings. Luck generally has a negative effect on the dismissal probability in columns (1) and (2), and columns (3) and (4) reveal that this effect is mainly driven by bad luck. In column (5), the interaction between *Performance Pressure* and *Luck* is negative and significant. Similar to our main results, in column (6), the positive effect of *Bad Luck* on the probability of dismissal is accentuated when principals face *Performance Pressure*. Thus, the interpretation of our results appears not to be sensitive to our estimation strategy.

Insert Table 3 about here

----

[20] We account for each separate coach-team dyad. For example, Zinédine Zidane was appointed head coach of Real Madrid from 01/2016 to 05/2018 and then again from 03/2019 to 06/2021, which enters our analysis as two separate coach-team dyads.

Since our specifications could be sensitive to our operationalization of performance pressure, we re-estimate equation (5) by relying on three alternative pressure measurements in Table 4. The variable *Performance Pressure Historical* puts a greater weight on past performance outcomes to build preseason expectations by setting alpha equal to 0.75 to calculate the weighted exponential moving average.[21] We further capture a team's preseason expectations by relying on Elo ratings from the website www.clubelo.com. Specifically, we build a league table rank based on preseason Elo ratings, and deviations from these expectations are captured by the variable *Performance Pressure Elo*. Both the *Performance Pressure Historical* and the *Performance Pressure Elo* metrics address potential concerns regarding the sensitivity of our results to how we measure teams' preseason expectations.

Finally, we also employ a measurement of performance pressure that captures the threat of relegation. Since relegation is accompanied by massive sporting and economic consequences (Dios Tena & Forrest, 2007; Moliterno et al., 2014), avoiding relegation is the key goal for many European football teams. Thus, falling in the relegation zone may represent the highest form of performance pressure in European football. To build the variable *Relegation Pressure,* we create a dummy variable that equals one if a team's league table standing is in the relegation zone and zero otherwise. By doing so, we treat the relegation playoff spots existing in certain leagues (i.e., Bundesliga and Ligue 1) as a direct relegation spot.

Table 4 shows that our results mostly remain robust to these alternative operationalizations of performance pressure. Columns (1) and (2) reveal that the interaction coefficients between *Luck* and both *Performance Pressure Historical* and *Performance Pressure Elo* are again negative and significant. Although the interaction coefficient in column (3) is negative and thus, in the proposed direction, we do not find a statistically significant moderating effect of *Relegation Pressure* on *Luck*.

[21] To derive the team's last season historical aspirations level, we proceed in two steps. First, we rely on www.transfermarkt.com to obtain historical end-season league table rankings. Then, we set the last season's HAL at zero the first time we have consecutive observations available. For example, RB Leipzig records the first historical end-season league table rank in 2009/2010. Thus, to calculate the HAL in the 2010/2011 season, we calculate 0.75 * $Rank_{t-1}$ + 0.25 * 0. HAL_{t-1}. If, for example, the final rank of a team who was playing in the third-tier league in a given season was five and the second-tier league has 20 teams and the first-tier league has 20 teams, a rank of 45 is assigned for that season. Moreover, the HAL always equals the highest possible rank in the top five league if a team's HAL is higher than the highest possible rank. Teams that were promoted last season receive a value of 20 (respectively 18 for the Bundesliga) for the $Rank_{t-1}$ variable.

However, columns (4)–(6) reveal that all interaction coefficients between the various pressure proxies and bad luck remain positive and significant. Again, we do not find support for the reinforcing role of performance pressure in instances of good luck. Overall, and among all alternative specifications, our conclusions remain qualitatively the same.

Insert Table 4 about here

----

5 CONCLUSION

Consistent with outcome bias, we find that club boards evaluate coaches based on luck. By differentiating between outcome-biased evaluations in instances of bad and good luck, our field evidence suggests that club boards penalize coaches for bad luck but do not reward them for good luck. Most importantly, head coaches facing bad luck are more likely to be penalized for bad luck when club boards experience mounting performance pressure. Thus, our results show that performance pressure is a key contingency factor accentuating outcome bias, primarily driven by instances of bad luck.

Our findings support the idea that overlooking the impact of performance outcomes on performance pressure may underestimate the outcome bias in principal-agent settings. Thus, the extent to which principals fall prey to outcome bias may be greater than previously assumed since many high-stakes evaluations are conducted under performance pressure.

While we find support for performance pressure being a key contingency for bad luck, we do not find a significant effect for good luck. These findings could be attributed to an asymmetrical response of principals to agents' experience of bad luck compared to good luck, consistent with the long-documented negativity bias in attentional allocation (Smith et al., 2006), which states that negative and positive stimuli are of distinct intensity (Baumeister et al., 2001; Ito et al., 1998; G. Peeters & Czapinski, 1990). Since negative events cause stronger cognitive reactions than neutral or good events (Ito et al., 1998; Taylor, 1991), agents receive less credit for good outcomes than they receive blame for bad outcomes (Erkal et al., 2022).

However, no consensus exists on this matter in the broader literature. For example, König-Kersting et al. (2021) find that in the context of financial agency, principals' outcome-biased evaluations are more pronounced after good outcomes than after bad outcomes. Our finding may further appear puzzling in the broader literature reporting that CEOs are rewarded for luck, often reflected in their compensation (Bertrand & Mullainathan, 2001). Building on this, Amore and Schwenen (2022) conclude that conditional on changes in employment, lucky CEOs are rewarded with higher pay at new firms. Moreover, prior literature even highlight an asymmetry in pay-for-luck by emphasizing that CEOs are more rewarded with increased pay for good luck than they are penalized with a decrease in pay for bad luck (Garvey & Milbourn, 2006). Recently, however, Daniel et al. (2020) did not find any asymmetry in pay-for-luck, indicating that CEOs are similarly rewarded and punished for good and bad luck.

One possible explanation for these partly conflicting findings in the literature could be that rewards and punishment run through different channels. For instance, our dependent variable, the probability of managerial dismissal, may effectively capture punishment but represents only an imperfect proxy for reward since dismissals usually represent the ultimate sanction (Fama & Jensen, 1983). Similarly, compensation could serve as a suitable measurement for rewards while capturing punishment only partially. Support for this notion stems from the argument that if reward and punishment constitute two distinct categories, they are unlikely to be the opposite of each other (Fiorillo, 2013).

Our findings also have crucial implications in our institutional setting. Particularly in the wake of performance pressure, a scenario in which subsequent performance improvements are urgently needed, the costs of replacing an unlucky but skilled coach before contract expirations seem substantial. Since only dismissals after actual bad performance on the pitch increase subsequent performance compared to a control group of nondismissals with similar bad performance (Flepp & Franck, 2021), urgently needed subsequent performance improvements seem unrealistic if club boards replace an unlucky coach under performance pressure. Importantly, in the organizational context in which performance outcomes are noisier and less frequently observed than in European football, the documented effects might be even more pronounced. Our results thus might constitute the lower boundary of the true effect of outcome bias.

Despite its exogeneity, a key advantage of our conceptualization, our luck measurement is not without limitations. We acknowledge that key opponent injuries only capture a specific aspect of luck in European football. Brechot and Flepp (2020) outline that football is a low-scoring game in which winning or losing is sometimes decided by a single goal; thus, the game is heavily influenced by randomness. Therefore, further sources of luck that we do not capture certainly exist. Moreover, another shortcoming may be that, theoretically, a diminishing marginal effect of the number of key opponent player injuries could partly explain the weak effect of good luck. A further potential setback of our setting is that the actual decision-making process among decision-making principals (i.e., club boards) remains a black box. Since some individuals might be better at coping with performance pressure (Mitchell et al., 2019), some club boards might also have more experience dealing with performance pressure than others. However, given the available data, we are convinced that this concern is addressed in the best possible way by including team-season fixed effects.

Notwithstanding these potential weaknesses, by using an innovative approach to measure (bad and good) luck in a field setting, we elucidate the crucial influence of performance pressure on outcome-biased evaluations. Future research could expand on our methodological idea to explore other key contingencies accentuating outcome bias.

Table 1: Descriptive statistics and pairwise correlations of the main variables of interest

	Mean	Sd	Min	Max	1	2	3	4	5	6	7	8	9	10
1 Coach Dismissal	0.010	-	0	1	1									
2 Luck	0.176	1.556	-6.500	9.500	-0.022	1								
3 Bad Luck	0.436	0.826	0.000	6.500	0.025	-0.738	1							
4 Good Luck	0.612	1.099	0.000	9.500	-0.013	0.862	-0.294	1						
5 Performance Pressure	-0.626	5.183	-19	17	0.087	-0.070	0.031	-0.076	1					
6 Point Percentage	0.477	0.201	0.000	1.000	-0.085	0.038	-0.012	0.044	-0.340	1				
7 Tenure	723.180	803.279	0	7,894	-0.026	0.010	0.002	0.016	0.060	0.038	1			
8 Games Played	17.897	10.811	1	38	-0.009	0.025	0.218	0.199	-0.041	0.041	0.108	1		
9 Focal Key Injuries	1.562	5.536	-15.000	71.000	0.003	-0.142	0.167	-0.076	0.070	-0.009	0.050	0.134	1	
10 UEFA Period Opponent	0.201	-	0	1	0.008	-0.005	0.021	0.008	0.011	-0.045	-0.003	-0.054	-0.002	1

Notes: Descriptive statistics and pairwise correlations are calculated for 15,458 team observations. We do not report standard deviations for dummy variables.

Table 2: Main Results

	Coach Dismissal (1/0)					
	(1)	(2)	(3)	(4)	(5)	(6)
Luck	-0.00141***	-0.00176**			-0.00221**	
	(0.001)	(0.001)			(0.001)	
Bad Luck			0.00273**	0.00434**		0.00516***
			(0.001)	(0.002)		(0.002)
Good Luck			-0.00056	0.00002		0.00036
			(0.001)	(0.001)		(0.001)
Performance Pressure					0.00443***	0.00370***
					(0.001)	(0.001)
Performance Pressure x Luck					-0.00055***	
					(0.000)	
Performance Pressure x Bad Luck						0.00150***
						(0.000)
Performance Pressure x Good Luck						0.00016
						(0.000)
Point Percentage		-0.03649***		-0.03635***	0.03803***	0.03142***
		(0.007)		(0.007)	(0.009)	(0.009)
Tenure		0.00009*		0.00009**	0.00011**	0.00012***
		(0.000)		(0.000)	(0.000)	(0.000)
Games Played		0.00005		-0.00003	-0.00004	-0.00022
		(0.000)		(0.000)	(0.000)	(0.000)
Key Focal Injuries		-0.00006		-0.00009	-0.00009	-0.00007
		(0.000)		(0.000)	(0.000)	(0.000)
UEFA Period Opponent		0.00133		0.00116	0.00074	0.00049
		(0.002)		(0.002)	(0.002)	(0.002)
Observations	15,458	15,458	15,458	15,458	15,458	15,458
R-squared	0.001	0.073	0.001	0.073	0.079	0.081
Team-Season FE	No	Yes	No	Yes	Yes	Yes
Coach FE	No	Yes	No	Yes	Yes	Yes

Notes: OLS regressions with robust standard errors clustered on games in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Cox Proportional Hazard Models

	Coach Dismissal (1/0)					
	(1)	(2)	(3)	(4)	(5)	(6)
Luck	-0.23438***	-0.16476***			-0.10648*	
	(0.088)	(0.060)			(0.062)	
Bad Luck			0.50489***	0.30897***		0.22435**
			(0.076)	(0.094)		(0.111)
Good Luck			0.17878**	0.00042		0.01456
			(0.072)	(0.088)		(0.090)
Performance Pressure					0.15052***	0.14012***
					(0.017)	(0.018)
Performance Pressure x Luck					-0.02304**	
					(0.012)	
Performance Pressure x Bad Luck						0.03387*
						(0.017)
Performance Pressure x Good Luck						-0.00635
						(0.018)
Point Percentage		-3.20677***		-3.25457***	-1.86343***	-1.90016***
		(0.394)		(0.391)	(0.429)	(0.424)
Games Played		0.06749***		0.06032***	0.06338***	0.05614***
		(0.007)		(0.008)	(0.007)	(0.009)
Key Focal Injuries		0.00415		0.00384	-0.02008	-0.02501
		(0.013)		(0.013)	(0.016)	(0.017)
UEFA Period Opponent		0.79286***		0.76490***	0.70887***	0.68248***
		(0.213)		(0.214)	(0.214)	(0.215)
Log Pseudolikelihood	-753.0782	-689.2655	-739.9878	-687.2887	-645.1702	-643.1022
Wald x^2	7.12	149.01	46.08	158.56	236.44	272.69
Season FE	No	Yes	No	Yes	Yes	Yes

Notes: Cox proportional hazard models with right-censoring based on coach spell durations with robust standard errors clustered on coach-team dyads. Clustering the standard errors at the coach, team, or game level does not alter the results. We report coefficients instead of hazard ratios. The analysis is based on 15,458 observations and includes 387 potential subjects (coach-team dyads) at risk and 149 failures (dismissals).
*** p<0.01, ** p<0.05, * p<0.1

Table 4: Alternative Operationalization of Performance Pressure

	Coach Dismissal (1/0)					
	(1)	(2)	(3)	(4)	(5)	(6)
Luck	-0.00236**	-0.00221**	-0.00122*			
	(0.001)	(0.001)	(0.001)			
Bad Luck				0.00550***	0.00488***	0.00199
				(0.002)	(0.002)	(0.002)
Good Luck				0.00055	0.00032	-0.00076
				(0.001)	(0.001)	(0.001)
Performance Pressure Historical	0.00445***			0.00360***		
	(0.001)			(0.001)		
Performance Pressure Historical x Luck	-0.00055***					
	(0.000)					
Performance Pressure Historical x Bad Luck				0.00167***		
				(0.000)		
Performance Pressure Historical x Good Luck				0.00026		
				(0.000)		
Performance Pressure Elo		0.00445***			0.00361***	
		(0.001)			(0.001)	
Performance Pressure Elo x Luck		-0.00057***				
		(0.000)				
Performance Pressure Elo x Bad Luck					0.00170***	
					(0.000)	
Performance Pressure Elo x Good Luxk					0.00019	
					(0.000)	
Relegation Pressure			0.03951***			0.02566***
			(0.007)			(0.007)
Relegation Pressure x Luck			-0.00428			
			(0.004)			
Relegation Pressure x Bad Luck						0.02147***
						(0.008)
Relegation Pressure x Good luck						0.00668
						(0.005)

Table 4: Alternative Operationalization of Performance Pressure (continued)

Observations	15,458	15,458	15,458	15,458	15,458	15,458
R-squared	0.079	0.079	0.080	0.081	0.081	0.082
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Team-Season FE	Yes	Yes	Yes	Yes	Yes	Yes
Coach FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: OLS regressions with robust standard errors clustered on games in parentheses
*** p<0.01, ** p<0.05, * p<0.1

REFERENCES

Adams, R. B., Hermalin, B. E., & Weisbach, M. S. (2010). The role of boards of directors in corporate governance: A conceptual framework and survey. *Journal of Economic Literature, 48*(1), 58–107.

Allison, S. T., Mackie, D. M., & Messick, D. M. (1996). Outcome biases in social perception: Implications for dispositional inference, attitude change, stereotyping, and social behavior. *Advances in Experimental Social Psychology*, *28,* 53–93. https://doi.org/10.1016/S0065-2601(08)60236-1

Amore, M. D., & Schwenen, S. (2022). Hiring Lucky CEOs. *The Journal of Law, Economics, and Organization.* Advance online publication. https://doi.org/10.1093/jleo/ewac021

Baron, J., & Hershey, J. C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569–579. https://doi.org/10.1037/0022-3514.54.4.569

Baumeister, R. F. (1984). Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, *46*(3), 610–620. https://doi.org/10.1037/0022-3514.46.3.610

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, *5*(4), 323–370. https://doi.org/10.1037/1089-2680.5.4.323

Baumeister, R. F., & Showers, C. J. (1986). A review of paradoxical performance effects: Choking under pressure in sports and mental tests. *European Journal of Social Psychology*, *16*(4), 361–383. https://doi.org/10.1002/ejsp.2420160405

Bertrand, M., & Mullainathan, S. (2001). Are CEOs rewarded for luck? The ones without principals are. *The Quarterly Journal of Economics*, *116*(3), 901–932. https://doi.org/10.1162/00335530152466269

Besters, L. M., van Ours, J. C., & van Tuijl, M. A. (2016). Effectiveness of in-season manager changes in English Premier League Football. *De Economist*, *164*, 335–356. https://doi.org/10.1007/s10645-016-9277-0

Böheim, R., Grübl, D., & Lackner, M. (2019). Choking under pressure–Evidence of the causal effect of audience size on performance. *Journal of Economic Behavior & Organization*, *168,* 76–93. https://doi.org/10.1016/j.jebo.2019.10.001

Brechot, M., & Flepp, R. (2020). Dealing with randomness in match outcomes: how to rethink performance evaluation in European club football using expected goals. *Journal of Sports Economics*, *21*(4), 335–362. https://doi.org/10.1177/1527002519897962

Brownback, A., & Kuhn, M. A. (2019). Understanding outcome bias. *Games and Economic Behavior*, *117*, 342–360. https://doi.org/10.1016/j.geb.2019.07.003

Bryson, A., Buraimo, B., Farnell, A., & Simmons, R. (2021). Time to go? Head coach quits and dismissals in professional football. *De Economist*, *169*, 81–105. https://doi.org/10.1007/s10645-020-09377-8

Budescu, D. V., & Bruderman, M. (1995). The relationship between the illusion of control and the desirability bias. *Journal of Behavioral Decision Making*, *8*(2), 109–125. https://doi.org/10.1002/bdm.3960080204

Byrne, K. A., Silasi-Mansat, C. D., & Worthy, D. A. (2015). Who chokes under pressure? The Big Five personality traits and decision-making under pressure. *Personality and Individual Differences*, *74*, 22–28. https://doi.org/10.1016/j.paid.2014.10.009

Daniel, N. D., Li, Y., & Naveen, L. (2020). Symmetry in pay for luck. *The Review of Financial Studies*, *33*(7), 3174–3204. https://doi.org/10.1093/rfs/hhz057

de Rond, M. (2014). The structure of serendipity. *Culture and Organization*, *20*(5), 342–358. https://doi.org/10.1080/14759551.2014.967451

Denrell, J., Fang, C., & Liu, C. (2019). In search of behavioral opportunities from misattributions of luck. *Academy of Management Review*, *44*(4), 896–915. https://doi.org/10.5465/amr.2017.0239

Dios Tena, J. de, & Forrest, D. (2007). Within-season dismissal of football coaches: statistical analysis of causes and consequences. *European Journal of Operational Research*, *181*(1), 362–373. https://doi.org/10.1016/j.ejor.2006.05.024

Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, *65*, 636–653. https://doi.org/10.1016/j.jebo.2005.12.004

Dreu, C. K. W. de (2003). Time pressure and closing of the mind in negotiation. *Organizational Behavior and Human Decision Processes*, *91*(2), 280–295. https://doi.org/10.1016/S0749-5978(03)00022-0

Eisenberger, R., & Aselage, J. (2009). Incremental effects of reward on experienced performance pressure: Positive outcomes for intrinsic interest and creativity. *Journal of Organizational Behavior*, *30*(1), 95–117. https://doi.org/10.1002/job.543

Ellis, A. P. J. (2006). System breakdown: The role of mental models and transactive memory in the relationship between acute stress and team performance. *Academy of Management Journal*, *49*(3), 576–589. https://doi.org/10.5465/amj.2006.21794674

Erkal, N., Gangadharan, L., & Koh, B. H. (2022). By chance or by choice? Biased attribution of others' outcomes when social preferences matter. *Experimental Economics*, *25*, 413–443. https://doi.org/10.1007/s10683-021-09731-w

Fama, E. F., & Jensen, M. C. (1983). Agency problems and residual claims. *Journal of Law and Economics*, *26*(2), 327–349. https://doi.org/10.1086/467038

Farrell, K. A., & Whidbee, D. A. (2003). Impact of firm performance expectations on CEO turnover and replacement decisions. *Journal of Accounting and Economics*, *36*, 165–196. https://doi.org/10.1016/j.jacceco.2003.09.001

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17. https://doi.org/10.1002/(SICI)1099-0771(200001/03)13:1<1::AID-BDM333>3.0.CO;2-S

Fiorillo, C. D. (2013). Two dimensions of value: dopamine neurons represent reward but not aversiveness. *Science*, *341*(6145), 546–549. https://doi.org/10.1126/science.1238699

Fisman, R. J., Khurana, R., Rhodes-Kropf, M., & Yim, S. (2014). Governance and CEO turnover: do something or do the right thing? *Management Science*, *60*(2), 319–337. https://doi.org/10.1287/mnsc.2013.1759

Flepp, R., & Franck, E. (2021). The performance effects of wise and unwise managerial dismissals. *Economic Inquiry*, *59*(1), 186–198. https://doi.org/10.1111/ecin.12924

Flepp, R., & Meier, P. F. (2024). Struck by Luck: Noisy Capability Cues and CEO Dismissal. *Working Paper*. https://ssrn.com/abstract=3904056

Flepp, R., Merz, O., & Franck, E. (2023). When the league table lies: does outcome bias lead to informationally inefficient markets? *Economic Inquiry*, *62*(1), 414–429. https://doi.org/10.1111/ecin.13163

Gardner, H. K. (2012). Performance pressure as a double-edged sword: Enhancing team motivation but undermining the use of team knowledge. *Administrative Science Quarterly*, *57*(1), 1–46. https://doi.org/10.1177/0001839212446454

Garvey, G. T., & Milbourn, T. T. (2006). Asymmetric benchmarking in compensation: executives are rewarded for good luck but not penalized for bad. *Journal of Financial Economics*, *82*(1), 197–225. https://doi.org/10.1016/j.jfineco.2004.01.006

Gauriot, R., & Page, L. (2019). Fooled by performance randomness: overrewarding luck. *Review of Economics and Statistics*, *101*(4), 658–666. https://doi.org/10.1162/rest_a_00783

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. https://doi.org/10.1037/0022-3514.60.4.509

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*(1), 42. https://doi.org/10.1037/1076-898X.10.1.42

Gurdal, M. Y., Miller, J. B., & Rustichini, A. (2013). Why blame? *Journal of Political Economy*, *121*(6), 1205–1247. https://doi.org/10.1086/674409

Harb-Wu, K., & Krumer, A. (2019). Choking under pressure in front of a supportive audience: Evidence from professional biathlon. *Journal of Economic Behavior & Organization*, *166*, 246–262. https://doi.org/10.1016/j.jebo.2019.09.001

Hilger, S., Mankel, S., & Richter, A. (2013). The use and effectiveness of top executive dismissal. *The Leadership Quarterly*, *24*(1), 9–28.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, *28*(4), 331–343. https://doi.org/10.1002/smj.582

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, *10*, 74–91. https://doi.org/10.2307/3003320

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology*, *75*(4), 887–900. https://doi.org/10.1037/0022-3514.75.4.887

Jedelhauser, F., Flepp, R., & Franck, E. (2023). Overshadowed by popularity: the value of second-tier stars in European football. *Journal of Sports Economics*, *24*(8), 1026–1054. https://doi.org/10.1177/15270025231187880

Jenter, D., & Kanaan, F. (2015). CEO turnover and relative performance evaluation. *Journal of Finance*, *70*(5), 2155–2184. https://doi.org/10.1111/jofi.12282

Kahn, L. M. (2000). The sports business as a labor market laboratory. *Journal of Economic Perspectives*, *14*(3), 75–94. https://doi.org/10.1257/jep.14.3.75

Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475. https://doi.org/10.1257/000282803322655392

Kahneman, D. (2011). *Thinking, fast and slow*. Penguin.

Kaplan, S. M. (2022). Putting a price on popularity: Evidence from superstars in the National Basketball Association. *Economic Inquiry*, *60*(3), 1357–1381. https://doi.org/10.1111/ecin.13065

Kausel, E. E., Ventura, S., & Rodríguez, A. (2019). Outcome bias in subjective ratings of performance: evidence from the (football) field. *Journal of Economic Psychology*, *75*, Article 102132. https://doi.org/10.1016/j.joep.2018.12.006

König-Kersting, C., Pollmann, M., Potters, J., & Trautmann, S. T. (2021). Good decision vs. good results: outcome bias in the evaluation of financial agents. *Theory and Decision*, *90*(1), 31–61. https://doi.org/10.1007/s11238-020-09773-1

Lefgren, L., Platt, B., & Price, J. (2015). Sticking with what (barely) worked: a test of outcome bias. *Management Science*, *61*(5), 1121–1136. https://doi.org/10.1287/mnsc.2014.1966

Lefgren, L., Platt, B., Price, J., & Higbee, S. (2019). Outcome based accountability: theory and evidence. *Journal of Economic Behavior & Organization*, *160*, 121–137. https://doi.org/10.1016/j.jebo.2019.02.012

Lewis, B. P., & Linder, D. E. (1997). Thinking about choking? Attentional processes and paradoxical performance. *Personality and Social Psychology Bulletin*, *23*(9), 937–944. https://doi.org/10.1177/0146167297239003

Liu, C., & de Rond, M. (2016). Good night, and good luck: perspectives on luck in management scholarship. *Academy of Management Annals*, *10*(1), 409–451. https://doi.org/10.1080/19416520.2016.1120971

Marshall, G. W., & Mowen, J. C. (1993). An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management*, *13*(3), 31–47. https://doi.org/10.1080/08853134.1993.10753956

Meier, P. F., Flepp, R., & Franck, E. (2023). Replication: Do Coaches Stick with What Barely Worked? Evidence of Outcome Bias in Professional Sports. *Journal of Economic Psychology*, *99*, Article 102664. https://doi.org/10.1016/j.joep.2023.102664

Meier, P. F., Flepp, R., Meier, P., & Franck, E. (2022). Outcome bias in self-evaluations: quasi-experimental field evidence from Swiss driving license exams. *Journal of Economic Behavior & Organization*, *201*, 292–309. https://doi.org/10.1016/j.jebo.2022.07.013

Mitchell, M. S., Greenbaum, R. L., Vogel, R. M., Mawritz, M. B., & Keating, D. J. (2019). Can you handle the pressure? The effect of performance pressure on stress appraisals, self-regulation, and behavior. *Academy of Management Journal*, *62*(2), 531–552. https://doi.org/10.5465/amj.2016.0646

Moliterno, T. P., Beck, N., Beckman, C. M., & Meyer, M. (2014). Knowing your place: Social performance feedback in good times and bad times. *Organization Science*, *25*(6), 1684–1702. https://doi.org/10.1287/orsc.2014.0923

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: the distinction between affective and informational negativity effects. *European Review of Social Psychology*, *1*(1), 33–60. https://doi.org/10.1080/14792779108401856

Peeters, T. (2018). Testing the wisdom of crowds in the field: transfermarkt valuations and international soccer results. *International Journal of Forecasting*, *34*(1), 17–29. https://doi.org/10.1016/j.ijforecast.2017.08.002

Pieper, J., Nüesch, S., & Franck, E. (2014). How performance expectations affect managerial replacement decisions. *Schmalenbach Business Review*, *66*, 5–23. https://doi.org/10.1007/BF03396867

Puffer, S. M., & Weintrop, J. B. (1991). Corporate performance and CEO turnover: the role of performance expectations. *Administrative Science Quarterly*, *36*, 1–19. https://doi.org/10.2307/2393427

Rubin, J., & Sheremeta, R. (2016). Principal–agent settings with random shocks. *Management Science*, *62*(4), 985–999. https://doi.org/10.1287/mnsc.2015.2177

Semmelroth, D. (2022). Time to say goodbye: A duration analysis of the determinants of coach dismissals and quits in major league soccer. *Journal of Sports Economics*, *23*(1), 95–120. https://doi.org/10.1177/15270025211034820

Smith, N. K., Larsen, J. T., Chartrand, T. L., Cacioppo, J. T., Katafiasz, H. A., & Moran, K. E. (2006). Being bad isn't always good: affective context moderates the attention bias toward negative information. *Journal of Personality and Social Psychology*, *90*(2), 210. https://doi.org/10.1037/0022-3514.90.2.210

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A user's guide to debiasing. *The Wiley Blackwell Handbook of Judgment and Decision Making*, *2*, 924–951. https://doi.org/10.1002/9781118468333.ch33

Spoelma, T. M. (2022). Counteracting the effects of performance pressure on cheating: A self-affirmation approach. *Journal of Applied Psychology*, *107*(10), 1804–1823. https://doi.org/10.1037/apl0000986

Staw, B. M., Sandelands, L. E., & Dutton, J. E. (1981). Threat rigidity effects in organizational behavior: A multilevel analysis. *Administrative Science Quarterly*, *26*(4), 501–524. https://doi.org/10.2307/2392337

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychological Bulletin*, *110*(1), 67–85. https://doi.org/10.1037/0033-2909.110.1.67

van Ours, J. C., & van Tuijl, M. A. (2016). In-season head-coach dismissals and the performance of professional football teams. *Economic Inquiry*, *54*(1), 591–604. https://doi.org/10.1111/ecin.12280

Wine, J. (1971). Test anxiety and direction of attention. *Psychological Bulletin*, *76*(2), 92–104. https://doi.org/10.1037/h0031332

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Wulff, J. N., Sajons, G. B., Pogrebna, G., Lonati, S., Bastardoz, N., Banks, G. C., & Antonakis, J. (2023). Common methodological mistakes. *The Leadership Quarterly*, *34*(1), Article 101677. https://doi.org/10.1016/j.leaqua.2023.101677

Wunderlich, F., Seck, A., & Memmert, D. (2021). The influence of randomness on goals in football decreases over time. An empirical analysis of randomness involved in goal scoring in the English Premier League. *Journal of Sports Sciences*, *39*(20), 2322–2337. https://doi.org/10.1080/02640414.2021.1930685

APPENDIX

Table A1: Real data example illustrating the construction of the luck, bad luck, and good luck variables

Team	*Season*	*Games played*	*Key opponent injuries*	*Cum. key opponent injuries*	*Median cum. key opponent injuries*	*Luck*	*Bad Luck*	*Good Luck*
Arsenal	18/19	1	0	0	0	0	0	0
Arsenal	18/19	2	0	0	0	0	0	0
Arsenal	18/19	3	0	0	0	0	0	0
Arsenal	18/19	4	0	0	0	0	0	0
Arsenal	18/19	5	1	1	1	0	0	0
Arsenal	18/19	6	1	2	2	0	0	0
Arsenal	18/19	7	0	2	2	0	0	0
Arsenal	18/19	8	0	2	2	0	0	0
Arsenal	18/19	9	0	2	2	0	0	0
Arsenal	18/19	10	0	2	2	0	0	0
Arsenal	18/19	11	0	2	2.5	-0.5	0.5	0
Arsenal								
Arsenal	18/19	16	0	4	5	-1	1	0
Arsenal	18/19	17	1	5	5	0	0	0
Arsenal	18/19	18	0	5	6	-1	1	0
Arsenal	18/19	19	0	5	6.5	-1.5	1.5	0
Arsenal	18/19	20	0	5	6.5	-1.5	1.5	0
…								
Arsenal	18/19	27	0	8	9	-1	1	0
Arsenal	18/19	28	3	11	9	2	0	2
Arsenal	18/19	29	1	12	10.5	1.5	0	1.5
Arsenal	18/19	30	0	12	10.5	1.5	0	1.5
…								
Arsenal	18/19	38	0	17	16	1	0	1

Notes: This is a real data extract that shows the development of luck, bad luck, and good luck for Arsenal during the 2018/19 season.

Table A2: Validity of key opponent injuries and their accumulation as a proxy for exogenous luck

	Goal Difference	*Points*	*Rank*	*Rank*
	(1)	(2)	(3)	(4)
Key Opponent Injuries	0.11759***	0.09178***		
	(0.026)	(0.019)		
Luck			-0.10077***	
			(0.019)	
Bad Luck				0.14792***
				(0.033)
Good Luck				-0.06601**
				(0.027)
Observations	18,058	18,058	18,058	18,058
R-squared	0.149	0.125	0.808	0.808
Team-Season FE	Yes	Yes	Yes	Yes

Notes: OLS regressions with robust standard errors clustered on games in parentheses.

*** p<0.01, ** p<0.05, * p<0.1