

Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure

Lotte Bransen^{†*} – Pieter Robberechts^{‡*} – Jan Van Haaren[†] – Jesse Davis[‡]

[†]SciSports, The Netherlands – [‡]KU Leuven, Belgium

*Lotte Bransen and Pieter Robberechts contributed equally to this paper.

Other Sports Track
#13582

1. Introduction

An iconic recent moment in the English Premier League was the final day of the 2011/2012 season, where Manchester City needed a win to secure the title but trailed Queens Park Rangers by one goal heading into injury time. At the point, the key question was how the players would respond. Would they up their game or wilt under the mental pressure? What happened next has cemented itself into the lore of Manchester City's: Edin Džeko equalized before Sergio Agüero scored to improbably secure Manchester City's first league title in 44 years. While not all game situations are as pressure packed as this one, soccer players are confronted with numerous situations that impose a high level of mental pressure.

While most existing soccer performance metrics focus on a player's technical and physical performances (e.g., [5, 14, 20, 21]), they typically ignore the mental pressure under which these performances were delivered. Yet, mental pressure is a recurrent concept in the analysis of a player's or a team's performance.^{1,2,3} Hence, a metric that quantifies how mental pressure affects the performance of soccer players would have four important use cases for soccer clubs.

1. Player acquisitions: A soccer club would clearly prefer having players who perform well under mental pressure. Being able to more accurately assess this characteristic for transfer targets would yield an additional insight that could guide player acquisition. *Example:* We identify Houssem Aouar and Xherdan Shaqiri as suitable replacements for Leicester City's star Riyad Mahrez.

2. Training: If a soccer club knows that a player consistently makes poor decisions in certain tense circumstances, this could be addressed. For example, the manager could coach the player on what to do in specific contexts. Furthermore, it may be possible to design training sessions tailored towards addressing these weaknesses. *Example:* We identify a large number of needless fouls under pressure as one of the fixable weaknesses of Orlando City's striker Dom Dwyer.

3. Tactical decisions: Certain actions are relatively more valuable or likely to succeed in high-pressure situations during soccer matches. Knowing these actions could help a soccer manager in

¹ <https://www.skysports.com/football/news/11667/11528605/juan-mata-is-still-not-trusted-by-jose-mourinho-in-the-big-games>

² <https://sporza.be/nl/2018/10/05/kijk-om-14-30u-naar-de-wekelijkse-persbabbel-van-hein-vanhaezebr/>

³ <https://www.theguardian.com/football/blog/2017/feb/02/arsenal-chelsea-watford-arsene-wenger-pressure>



his tactical planning for a match. *Example:* We identify crossing as a strategy with underused potential in high-pressure situations for Manchester United.

4. Lineups and substitutions: Knowing which players on a club perform well under mental pressure could be used to help inform a manager's decision making in terms of which players to line up in anticipation of a crucial game and which players to substitute on or off when a game gets tense. *Example:* While Juventus' central defenders Benatia, Bonucci, Chiellini and Rugani are of equal strength in normal game situations, we identify Benatia and Bonucci as the best central duo in high-pressure situations.

This paper takes a first step towards objectively providing insight into the question: How will a soccer player perform and behave during high-mental-pressure game situations? To explore this question, we employ the following approach:

1. For each situation in a soccer game, we develop a machine learned model to estimate how much mental pressure the player possessing the ball experienced using a combination of match context features (e.g., whether the game is a rivalry and the gap between the teams' current and desired league positions) and the current game state (e.g., the score and time left).
2. Since mental pressure could affect a player's performance in different ways such as his decision making (i.e., how he selects an action from several possible choices) or how well he executes a chosen action, we develop machine learned models to evaluate three aspects of each action: the choice of action, the execution of the chosen action, and the action's expected contribution to the scoreline.
3. To assess how a player reacts to mental pressure, we compare his performance metrics across different levels of mental pressure.

Using this methodology, we analyzed event data for 6,858 matches from 7 leagues. Unlike tracking data, event data is widely available for a variety of leagues across the world. As a result, teams can use our metric not only to better evaluate their own players' contributions in the crucial moments of a game, but also for recruiting players and for comparing a team's own players with players in other teams and leagues.

2. Measuring Mental Pressure

We hypothesize that high mental pressure arises in situations that may significantly impact a team's ability to achieve its goals. Specifically, two factors affect the pressure level:

1. The context surrounding the match affects the pre-game pressure. For example, a rivalry game or a game directly impacting relegation will be more tense than a typical end-of-season game with little to nothing at stake.
2. The events in the game itself affect the in-game pressure level. Pressure mounts in close games, particularly as time winds down because a goal would increase the chances of a favorable outcome. Conversely, pressure decreases when the margin is big as a goal would only have a small impact on the expected outcome.

Therefore, we develop two novel metrics: one captures the pre-game pressure and the other the in-game pressure. Concretely, for a game g currently at game state xt , the total pressure is given as a combination of these two metrics:

$$Pressure(g, xt) = Pressure_{pre-game}(g) + Pressure_{in-game}(xt)$$

Next, we describe how we compute each metric.

2.1. Pre-game mental pressure

Ideally, each game would have a category denoting its pressure level (e.g., no pressure, low pressure, normal pressure, high pressure). However, no such labeling exists and it may be difficult to define such categories with enough precision to enable manually labeling the data. We exploit the fact that there is an ordering relationship between different pressure levels. Moreover, given two matches, assessing which match has higher stakes is easier than assigning a pressure level category to each match [18]. Therefore, we collect such judgments from soccer experts and use them to train a machine-learned ranker. The ranker learns to assign scores to pairs of games, such that a higher score is assigned to the game with the higher pre-game pressure. These scores define our pressure metric.

Figure 1 shows Everton's pre-game pressure level for each league match in the 2017/2018 season. Pressure starts to mount after the sixth game due to a string of poor performances that see the manager sacked after game nine. The pressure remains high as Everton hovers around the relegation zone. A string of good results sees the pressure abate as they climb the table. Towards the season's end, pressure remains moderate until they are ruled out of contention for the Europa League.

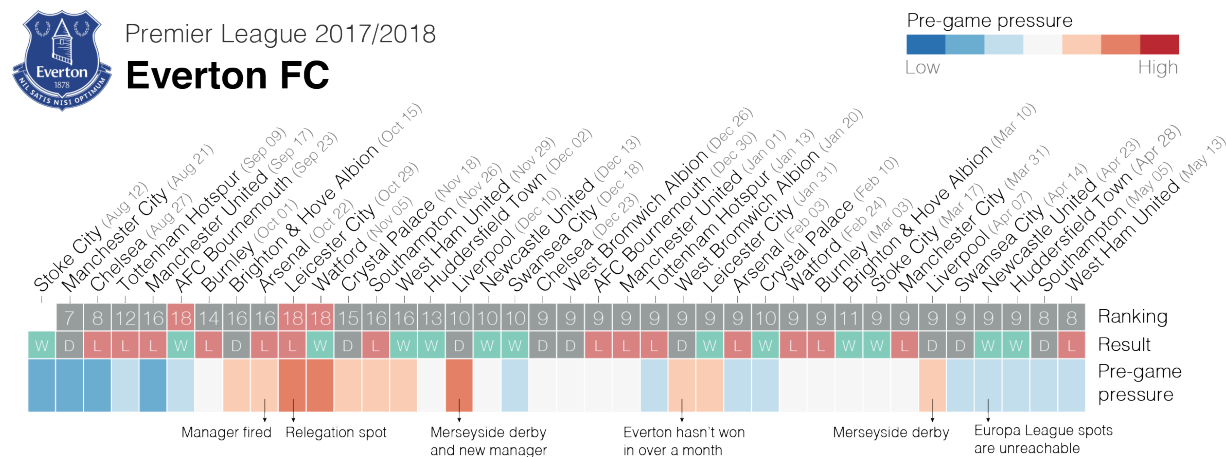


Figure 1. Pre-game pressure levels for Everton during the 2017/2018 Premier League season. The pressure mounts during Everton's poor start to the season and peaks when they end up on a relegation spot. Pressure decreases again once the new manager Allardyce has taken Everton to mid-table safety.

The model

The magnitude of the stakes and the importance of achieving success are important pressure facilitators [1]. Therefore, we need to construct features that capture this relationship. We consider the following four broad categories of features:

Team ambition. Each team will have ambitions for the season, such as winning the league or simply staying up, that affect its pre-game pressure level. We capture ambition by clustering the teams in each league into four groups using each team's result in previous seasons, transfer value of its top-20 players, spending on loans and Football Manager's reputation score, which reflects how prestigious a club is.

Game importance. Capturing how much a game will affect a team's chance to achieve its ambitions requires estimating how the current game's outcome will affect the probability that the team reaches a certain season outcome (e.g., avoiding relegation). We do this using an Elo-based model. Based on the pre-game league table, we simulate the rest of the season. Next, we use the Kendall-Stuart tau-c to measure the association between each possible game result (win-draw-loss) and each expected final league outcome (e.g., relegated, league champion) [13].

Recent performance. Soccer clubs are also subject to pressure based on recent form. Particularly for a big club, several consecutive poor performances will ratchet up the pressure. We capture this pressure source using the number of points obtained and the deviation from the expected performance using Elo ratings over the last five games.

Game context. Specific characteristics of a game will affect pressure, namely: game location (i.e., home or away), the rivalrousness of the opponent as determined by Football Manager data, the match attendance, and how long ago the coach was appointed.

We obtained pairwise rankings of games from a number of soccer experts and used them to fit a Gradient Boosted Ranking Trees model. Appendix A.1 provides details on the data and model.

2.2. In-game mental pressure

During the game, the teams' pressure levels will change over time based on the current game state. We argue that pressure should mount when scoring a goal increases the chance of a favorable match outcome, and subside when a goal would only have a small impact on the expected outcome. One way to estimate the impact of scoring or conceding a goal on the expected match outcome is to measure the difference in win probability between the current game state and the two hypothetical game states where the home or away team has scored an additional goal.

A team's in-game pressure level at time t is the sum of the increase in win probability and the increase in tie probability if the team would score at time t . Figure 2 shows the pressure levels and win probabilities throughout the nerve-wracking Everton vs Watford game in the 2017/2018 English Premier League season. Our pressure model reacts to events impacting the win probabilities such as goals. For instance, Everton's pressure level increases after Watford's opening goal and drops when Watford double their lead. However, Everton's pressure level sharply increases after their first goal and so does Watford's pressure level after the equalizer.

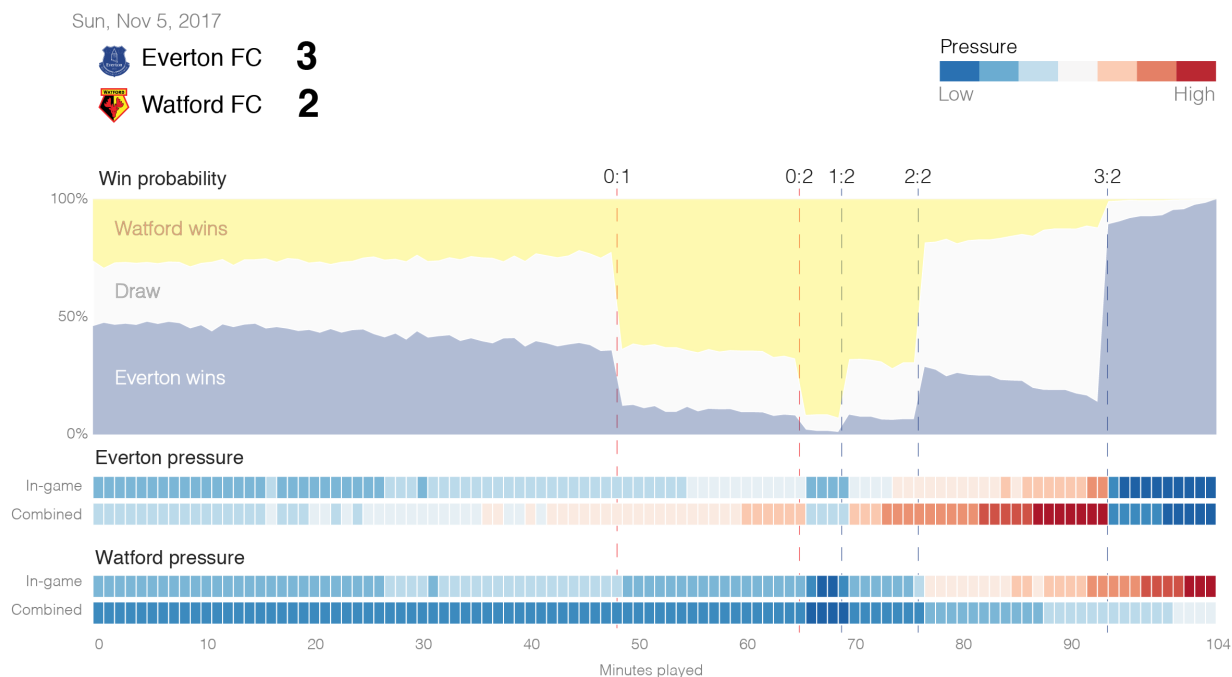


Figure 2. Evolving pressure levels (**bottom**) and win probabilities (**top**) in Everton's 3-2 win against Watford in the 2017/2018 English Premier League season. Pressure mounts when scoring increases the chance of a favorable match outcome, and subsides when a goal would only have a small impact on the expected outcome. At the start of the game, each team has a low pressure level since there is still enough time left to overcome the other team scoring and win the game.

The model

The low-scoring nature of soccer complicates building a win probability model. Therefore, instead of directly modeling the win-draw-loss probabilities, we predict the future number of goals that a team will score. By estimating the likelihood of each possible path to a win-draw-loss outcome, our model can capture the uncertainty of the win-draw-loss outcome in close games. Specifically, given the game state at time t , we model the probability distribution over the number of goals each team will score between time $t+1$ and the end of the match. From this, we can derive a distribution over the predicted final number of goals for each team as the sum of its number of goals at time t and predicted number of future goals after time t .

To deal with the variable duration of games due to stoppage time, we split each game into $T=100$ time frames, each corresponding to a percentage of the game. We model the predicted number of goals that the home ($y>t, home$) and away ($y>t, away$) team will score after time t , as independent Binomial distributions:

$$y>t, home \mid \theta t, home \sim B(T-t, \theta t, home),$$

$$y>t, away \mid \theta t, away \sim B(T-t, \theta t, away),$$



where the θ parameters represent each team's estimated scoring intensity in the t^{th} time frame. We estimate these scoring intensities from the current game state, which we describe using the following features: the number of goals scored, the goal difference, the number of yellow and red cards, the difference in Elo-ratings of the teams, the average number of attacking passes in the previous 10 time frames, and the average percentage of duels won in the previous 10 time frames. Since the importance of game state features varies over time in a non-linear way, we model the scoring intensities using a temporal stochastic process. This approach allows us to share information and to perform coherent inference between time frames.

We trained our in-game model using PyMC3's Auto-Differentiation Variational Inference (ADVI) algorithm [12] on data for the 2014/2015 through 2017/2018 seasons of the English Premier League, Spanish LaLiga, and German Bundesliga. A detailed description of the validation of this model can be found in Appendix A.2.

3. Measuring Player Performance

Pressure could affect players' performances in different ways. Hence, when a player performs an action in a match, there are arguably three important aspects of the action that should be evaluated:

1. **Total contribution:** How helpful was the result of an action in terms of increasing the team's chance of scoring or preventing the other team from scoring? For example, a successful through ball that puts a teammate 1-on-1 with the goalkeeper will have a high contribution.
2. **Quality of the decision:** Did the player take the best possible action? Even if an action helped his team, it is possible that another, even better action was possible. That is, the player could have made a better decision. For example, in the Tottenham - Manchester City match on October 29th, 2018, David Silva chose to pass instead of shoot when he was several meters in front of an open goal (see Figure 3). This was clearly a poor choice.
3. **Quality of the execution:** How well did the player perform the chosen action? A player may make the correct decision, such as shooting at an open goal, but simply execute the action poorly (e.g., sky the shot over the bar).

In recent years, several performance metrics have been introduced to capture the contributions of actions (e.g., [4, 7, 14]). However, to the best of our knowledge, no such performance metrics have been proposed yet for measuring the quality of a decision or an action's execution. Therefore, we use an existing metric to capture an action's contribution and introduce two novel metrics to evaluate decisions and executions. The following table gives some illustrative examples about what constitutes low, average, and high ratings according to each of our metrics.

Example	Contribution	Decision	Execution
A midfielder shoots from 35 meters out and scores, while a simple pass could have put a teammate 1-on-1 with the goalkeeper.	High	Low	High
The left back crosses the ball diagonally over 60 meters and reaches the right winger.	Average	Average	High

A midfielder attempts a through ball to put the striker in front of the opponent's goal. However, the through ball is inaccurate and the goalkeeper picks up the ball.	Low	High	Low
--	-----	------	-----

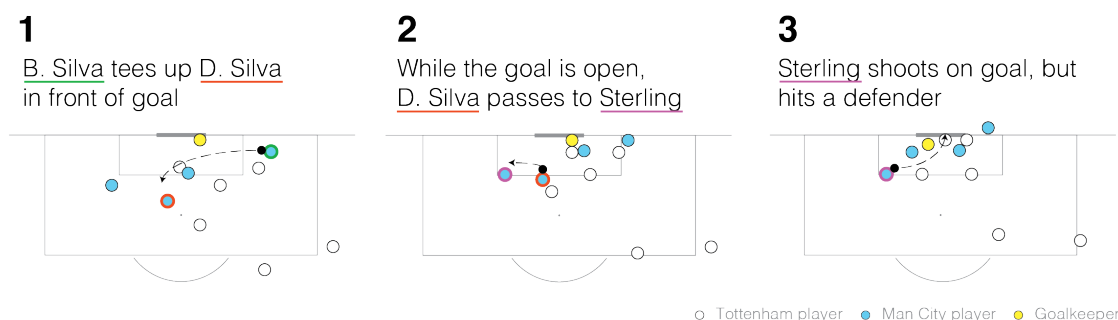


Figure 3. During the Tottenham - Manchester City match on October 29th, 2018, David Silva chose to pass to his teammate Raheem Sterling instead of to shoot when he was several meters in front of an open goal. This is the lowest-rated decision for the 2018/2019 Premier League season up until November 11th, 2018.

3.1. Contribution rating

As our contribution rating, we use the metric introduced by Decroos et al. [7]. It considers 22 different types of actions, such as a shot, pass, or dribble, all of which can either succeed or fail. In either case, the result of an action is that it modifies the game state. The goal of the contribution rating is to measure how valuable this resulting change of game state is. It does so by computing the difference between the game state values before and after an action, where the value of a game state reflects its likeliness of yielding a goal. Formally, the contribution rating for an action ai with outcome oi in a game state si is:

$$CR(si, ai, oi) = V(sj) - V(si),$$

where s_j represents the game state resulting from performing action ai with outcome oi in game state si , and $V(s)$ gives the value for a game state s .

The contribution of an action depends on several factors, including its type, location on the pitch, and outcome. Intuitively, the contribution rating is positive if the resulting game state has increased the team's chances of scoring (e.g., the ball moved to a more dangerous area via a successful dribble or pass) or decreased an opponent's chances of scoring (e.g., the keeper saved a shot). In contrast, the contribution rating is negative if the resulting game state has decreased the team's chances of scoring (e.g., a failed cross) or increased an opponent's chances of scoring (e.g., the opponent intercepted a pass).

3.2. Decision rating

In each game state, a player must decide which action to perform from several possibilities. To understand if a player selected a good action to perform, we need to (1) abstract away from the actual result of the action (i.e., we need to consider what happens both when the action succeeds and when it fails), and (2) consider how the chosen action relates to the other possible actions that the game state afforded. To this end, we measure the quality of a player's choice in a game state by comparing the expected contribution rating of the chosen action with the expected contribution rating across all possible actions in a game state. Formally, the decision rating for an action ai in a game state si corresponds to:

$$DR(si, ai) = ECR|si, ai - ECR|si.$$

The first term is simply the chosen action's expected contribution rating. The second term requires determining other possible actions in a game state, which is challenging because we use event data and thus do not know the precise locations of the players on the pitch. Therefore, we estimate the expected contribution rating across the possible actions by predicting the next action's contribution rating given the current game state.

Computing the expected contribution rating for the chosen action

Since an action can be successful or unsuccessful, we compute the expected value of its contribution rating as the weighted sum of the contribution of both outcomes:

$$ECR|si, ai = P(oi+) \cdot CR(si, ai, oi+) + P(oi-) \cdot CR(si, ai, oi-),$$

where $P(oi+)$ is the probability that action ai succeeds, and $P(oi-)$ is the probability that it fails.

To predict the probability that a given action will be successful, we train a binary classifier on historical match data for each type of action (e.g., pass, dribble or interception). We use a Gradient Boosted Trees model because it produces well-calibrated probability estimates (see Appendix A.3), which is important for this task. Positive examples are successful actions (e.g., passes finding a teammate or shots resulting in a goal), whereas negative examples are unsuccessful actions (e.g., missed shots). The features include an action's start and end locations, the body part used to execute the action, and the start and end locations of previous actions in the sequence. For action types for which the end location might contain information about the success (e.g., for a shot), we exclude this feature.

Estimating the expected contribution rating across all actions

We use historical observations of actions performed in highly similar game states to estimate the expected contribution rating across the possible actions in a game state. To this end, we train a Gradient Boosted Trees model that predicts the next action's contribution rating given the current game state. We use the following features to describe the current game state: the current location of the ball, the start and end locations of the previous two actions, the action types of the previous two actions, and the speed of the sequence (consisting of the last two actions) captured in distance traveled and time covered. A formal specification and evaluation of this model can be found in Appendix A.4.

3.3. Execution rating

The execution rating attempts to assess whether a player executed an action well or not, regardless of whether the selected action was a good choice. Intuitively, we want to reward players who successfully perform a difficult action such as completing a through ball or connecting on a long-range shot. Similarly, we want to punish players who flub an easy action such as having a lateral pass to an open teammate under no pressure. To this end, we measure the execution rating by computing the difference between the observed outcome of the action (e.g., did the cross reach a teammate or did the shot go in) and the predicted probability that the action would be successful. As an example, our model assigns a high execution rating to Zlatan Ibrahimović's first goal with the Galaxy against Los Angeles FC, which was a brilliant long-range shot over the keeper. Formally, the execution rating for an action ai with outcome oi corresponds to:

$$ER(ai, oi) = [oi+] - P(oi+),$$

where $[oi+]$ takes the value of one if oi succeeds and is zero otherwise, and $P(oi+)$ is given by the action success predictor from the previous section.

4. Measuring Player Performance Under Mental Pressure

To analyze the performances of players under different mental pressure levels, we need to combine our pressure model and three performance metrics. First, we use our pressure model to make a global ranking of the mental pressure level of each situation in all the analyzed matches. We then label each situation's pressure level as follows:

1. High-pressure situations fall in the top 20% of the ranking;
2. Normal-pressure situations fall in the middle 60% of the ranking; and
3. Low-pressure situations fall in the bottom 20% of the ranking.

Second, we rate all actions in all considered matches using each of our three performance metrics. Third, for a given player or team, we aggregate our three performance metrics as a function of the pressure level under which the action was performed. In our analysis, we consider the following aggregations of our performance metrics:

- Average contribution rating per 90 minutes;
- Average contribution rating per 90 minutes per action type;
- Percentile rank contribution rating per 90 minutes;
- Average decision rating;
- Average execution rating per action type;
- Percentile rank execution rating per action type.

Since a player's contribution heavily depends on his number of minutes played, we always report contribution ratings normalized per 90 minutes of play. Since the execution ratings for different types of actions have different distributions (i.e., a shot is typically less likely to result in a goal than a pass is to reach a teammate), we never aggregate execution ratings across action types.

5. Use Cases

Our analysis relies on event data from the English Premier League, Spanish LaLiga, German Bundesliga, Italian Serie A, French Ligue 1, Dutch Eredivisie, and North-American Major League Soccer (MLS). For each league, we consider the 2016/2017 and 2017/2018 seasons as well as the ongoing 2018/2019 seasons up until November 11th, 2018. Additionally, we include the 2014/2015 and 2015/2016 Premier League seasons to analyze the long-term evolution of players.

For each situation in the matches we analyzed, we compute the pressure level as well as all three of our performance metrics for the performed action. We compare a player's average contribution per 90 minutes, as well as his average decision, and execution ratings in low-pressure situations to those in high-pressure situations. We only consider players who played at least 900 minutes in total, including at least 180 minutes in each of high- and low-pressure situations.

In the remainder of this section, we demonstrate how our metrics can help clubs, managers and coaches in addressing the following key questions:

1. **Player acquisition:** Does a player perform well under pressure?
2. **Training:** Which are a player's recurring poor decisions in certain tense circumstances, such that they can be addressed during training?
3. **Tactical guidelines:** Which tactical plans are more likely to succeed in high-pressure situations?
4. **Lineups and substitutions:** Which players should a manager line up in anticipation of a crucial game or substitute when a game gets tense?

5.1. Player acquisition: Replacing Riyad Mahrez at Leicester City

In this use case, we step into the shoes of Leicester City's technical director in the summer of 2018. After granting Algerian winger Riyad Mahrez's wish to play for a top-six club, he must bring in a suitable replacement for his star player. Hence, this analysis only considers the data until the end of the 2017/2018 season.

As seen in Figure 4a, our analysis of Riyad Mahrez' contribution ratings per action type shows that the winger contributed mostly to Leicester City's performances with valuable dribbles, take-ons, shots, and crosses. Furthermore, our analysis of his 986 high-pressure minutes shows that he performs particularly well in tense situations. Namely, by both making better decisions and executing his shots better, Mahrez contributed considerably more to Leicester City's performances in high-pressure than in low-pressure situations.

Across Europe's top-five leagues, we identify three candidate replacements under the age of 28 who, like Riyad Mahrez, excel in dribbles, take-ons, shots and passes: Xherdan Shaqiri (Stoke City), Houssem Aouar (Olympique Lyonnais), and Rachid Ghezzal, who eventually joined Leicester City from Monaco on August 5th, 2018. Figure 4b shows the total contribution per 90 minutes in low- and high-pressure situations for these players. While not contributing as much as Mahrez, both Aouar's (547 high-pressure minutes) and Shaqiri's (1477 high-pressure minutes) contribution improve under pressure. In contrast, Leicester City's signing Ghezzal (358 high-pressure minutes) appears to choke under pressure. Figures 4c and 4d show how the average decision and shot execution ratings depend on the pressure level. Mahrez and Shaqiri improve their decision making when under high mental pressure, whereas only Mahrez and Aouar improve their execution

ratings. When under pressure, Ghezzal both seems to make worse decisions and has difficulty executing shots.



Finding a replacement for
R. Mahrez

Suggested alternatives:

- R. Ghezzal
- H. Aouar
- X. Shaqiri
- Average midfielder

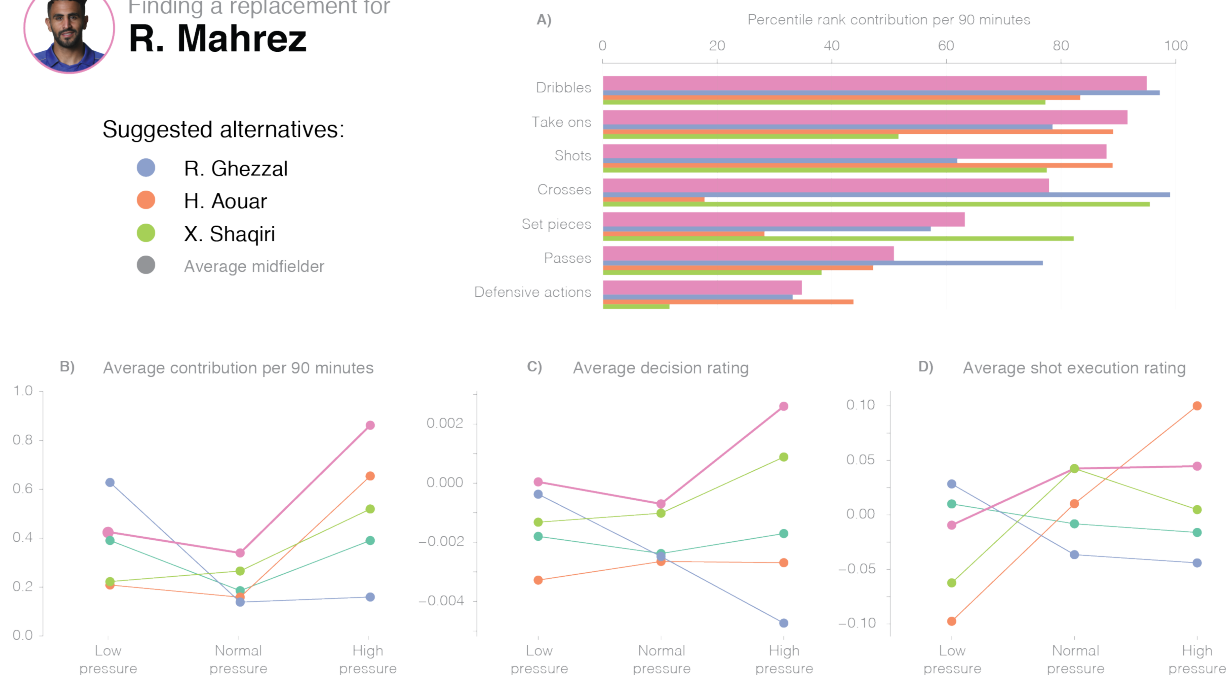


Figure 4. A comparison between Riyad Mahrez' and his suggested replacements at Leicester City: Aouar, Ghezzal and Shaqiri. Graph A compares the contribution ratings per action type for these players. Graphs B, C and D compare their performance under pressure.

We conclude that Ghezzal was probably not the best option for Leicester City to buy as a replacement for Mahrez. We observe that Mahrez was able to deliver in high-pressure situations whereas Ghezzal chokes under high mental pressure. Therefore, our analysis suggests that both Aouar and Shaqiri would have been better replacements for Mahrez as they have similar skills, including the ability to perform in high-pressure situations.

5.2. Training: Dom Dwyer's Fouls

If a club knows that a player consistently makes poor decisions in certain tense circumstances, this could be addressed. The manager could give a player insight in how his decisions have a negative impact on the team and coach the player on how to act differently. Furthermore, it may be possible to design training sessions tailored towards addressing these weaknesses. To showcase this use case, we compare the fouls committed by Orlando City's central striker Dom Dwyer in low-pressure (916 minutes) and high-pressure situations (2135 minutes). As Figure 5 illustrates, Dwyer commits significantly more fouls inside the opponent's penalty box when the pressure is high. Moreover, our metrics indicate that the majority of his fouls in high-pressure situations have a negative decision rating, meaning that they simply hand the possession to the opponent (instead of, for example, blocking a counter attack). His coach could use this insight to work on this aspect of Dwyer's game during training.



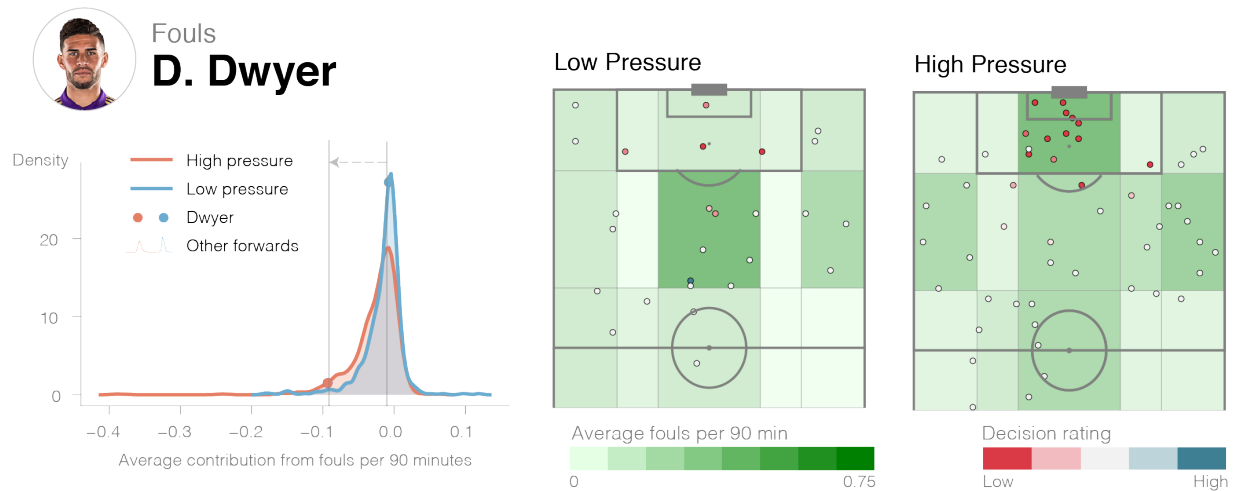


Figure 5. A comparison between the fouls committed by Dom Dwyer in low-pressure and high-pressure situations. While under pressure, Dwyer commits significantly more needless fouls, handing ball possession to the opponent.

5.3. Tactical planning

Manchester United: dribble or cross?

Manchester United relies on dribbling in high-pressure situations (1818 high-pressure minutes), with the number of dribbles per 90 minutes more than doubling from 19.41 to 43.88 under pressure. However, as Figure 6 illustrates, this is not a very successful strategy as the average contribution from dribbles per 90 minutes declines under pressure for all offensive players except Rashford. Our metrics suggest that crossing the ball would be a more successful strategy. Especially the crosses from players like Valencia, Pogba and Mata contribute to great scoring opportunities under pressure. Moreover, with players like Lukaku and Fellaini, United has the power to finish off these crosses.

Cross or Dribble

Manchester United

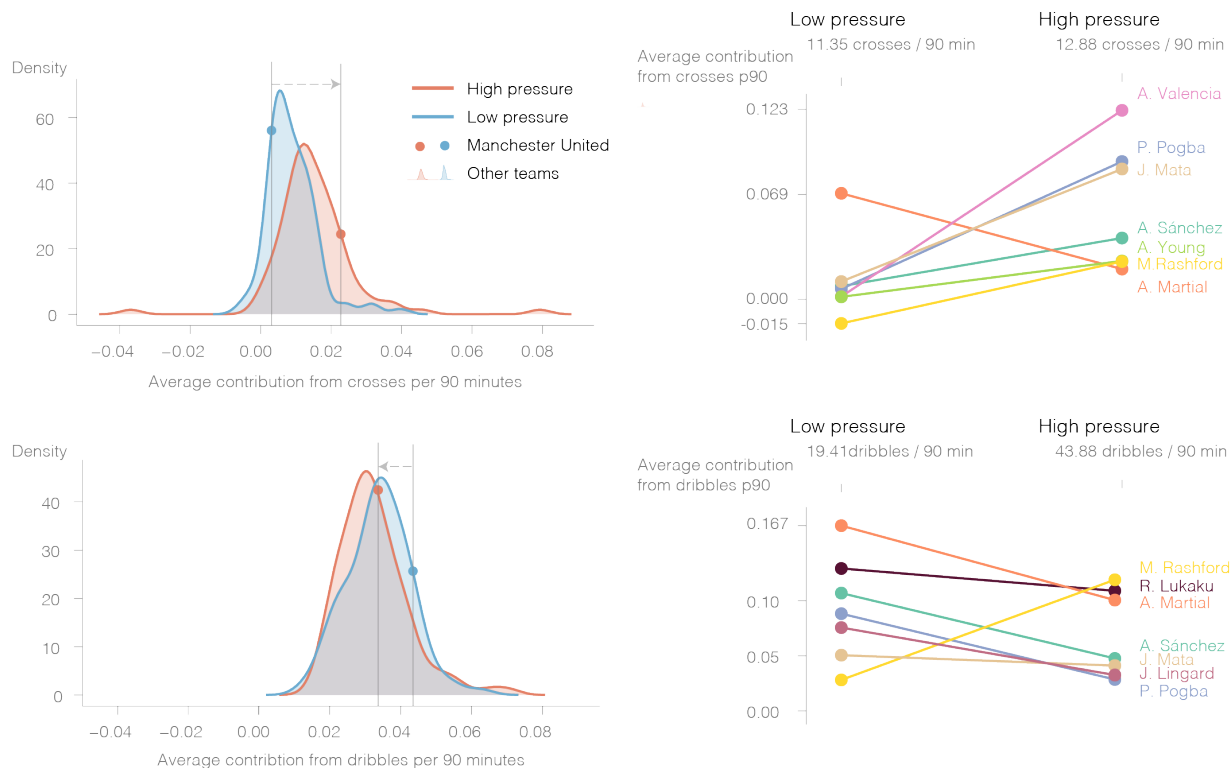


Figure 6. Manchester United typically attempts to use dribbles to force a breakthrough in high-pressure situations. Yet, the contribution from these dribbles decreases, while the contribution from crosses increases under pressure. This observation suggests that, in tense games, crossing the ball might be a more valuable strategy than dribbling.

Throw-ins: better decisions under high mental pressure

Throw-ins are the action type for which we observe the largest difference between low-pressure and high-pressure situations. For almost all teams, their contribution and decision ratings increase, while their execution ratings remain the same. This is an interesting observation which suggests that teams miss the opportunity to create more danger with throw-ins early in the game. Moreover, Liverpool's manager Jürgen Klopp also recognized the increasing importance of throw-ins in soccer by appointing a dedicated throw-in coach in the summer of 2018.⁴ As an example, Figure 7 shows the average decision ratings per 90 minutes of Leicester City's throw-ins and how the throw-in behavior differs between low-pressure and high-pressure situations. While one might assume that the lower decision ratings in low-pressure situations arise from more backward throws (e.g., to ensure ball possession), this is not the case. Instead, players are more inclined to throw forward in

⁴ <http://global.espn.com/soccer/club/liverpool/364/blog/post/3615088/liverpool-sign-throw-in-record-holder-why-jurgen-klopp-thinks-thomas-gronemark-can-be-key>

low-pressure situations. Throw-ins from a position near the opponent's goal line are the exception, with more balls thrown inside the penalty box in high-pressure situations. The Leicester City manager could use this information to train his players to apply their high-pressure throw-in behavior throughout the entire game. Similarly, the opposing manager could use these insights in his tactical discussion to highlight how his players should alter their defensive organization against throw-ins in different game situations when playing Leicester City.

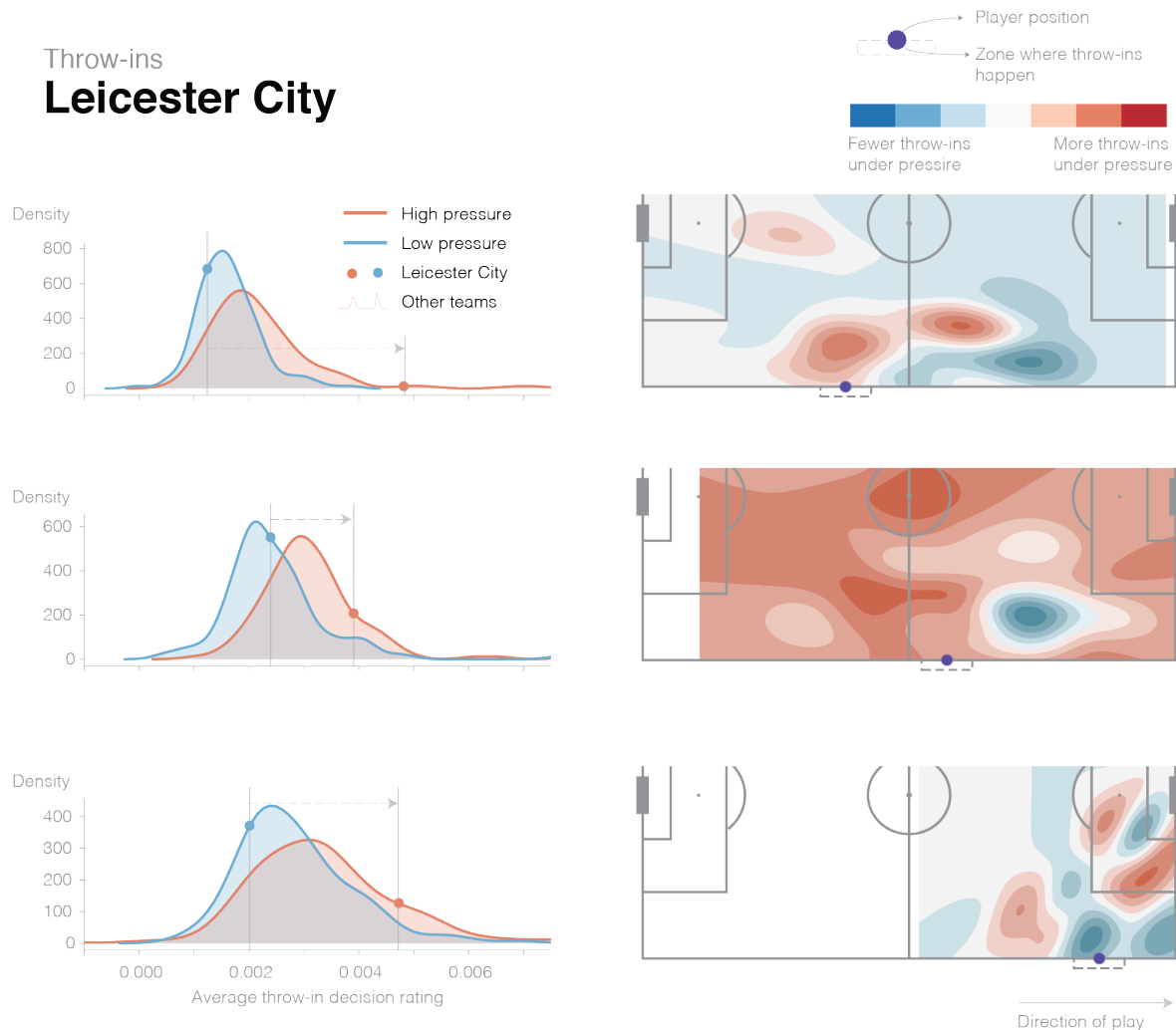


Figure 7. Average decision ratings per 90 minutes of Leicester City's throw-ins, compared to the league average (**left**) and probability maps of where these throw-ins end up (**right**). Data is modeled as a two-dimensional Gaussian distribution with red zones showing where Leicester is more likely to throw under pressure (compared to when the pressure is low) and blue zones where they are less likely to throw.

5.4. Lineups and substitutions

Sending the right players on the pitch is a key task of the manager. Essentially, this task comes down to predicting which players will perform best in a given situation. A player's expected performance under pressure is therefore an important variable while making these decisions. Our metrics could help inform a manager's decision making in terms of which players to line up in high-pressure games or which players to substitute during a tense game. As an example, Figure 8 compares the average contribution per 90 minutes of Juventus' central defenders Medhi Benatia (457 high-pressure minutes), Leonardo Bonucci (201 high-pressure minutes), Giorgio Chiellini (470 high-pressure minutes) and Daniele Rugani (278 high-pressure minutes), PSG's offensive midfielders Ángel Di María (451 high-pressure minutes) and Julian Draxler (359 high-pressure minutes), and Chelsea's strikers Olivier Giroud (617 high-pressure minutes at Arsenal and 127 at Chelsea) and Álvaro Morata (477 high-pressure minutes). For each of these cases, our metrics suggest that different players should be selected in high-pressure situations. While all of Juventus' central defenders have a similar average contribution in games with an average pressure, Benatia and Bonucci outperform Chiellini and Rugani in high-pressure situations. Similarly, di María levels up when the pressure mounts, outperforming Draxler. Finally, although Morata has in general a higher contribution than Giroud, he seems to choke under pressure; Giroud, on the other hand, established himself as a clutch-goal game-killing striker at Chelsea.⁵ Note, however, that we include data from Giroud's period at Arsenal due to lack of enough high-pressure moments at Chelsea.



Figure 8. A comparison of the average contribution per 90 minutes between Juventus' defenders, PSG's offensive midfielders and Chelsea's strikers. These graphs show that it might be worthwhile to select different players in high-pressure situations. For Giroud, we include data from his period at Arsenal due to lack of enough high-pressure moments at Chelsea.

⁵ <https://theprideoflondon.com/2018/10/01/chelsea-double-standard-alvaro-morata-olivier-giroud/>

6. Observations

This section presents some interesting observations which are based on the analysis of our performance and pressure metrics. First, we show that Neymar's performance declines under mental pressure. Second, we analyze the performance under pressure of three MLS talents. Third, we remark that Liverpool tends to buy players who perform better under high pressure. Finally, we show that our findings go beyond detecting performance changes due to fatigue.

6.1. Neymar's performances decline under mental pressure

Figure 9 shows how the performance of Neymar (Barcelona and PSG) varies between high- and low-pressure situations. Neymar evolves from a player with a very high average contribution per 90 minutes under low pressure (based on 997 low-pressure minutes) to an average player in high-pressure situations (based on 472 high-pressure minutes). Further analysis shows that this decrease in performance can be explained by a poor decision making behaviour. Neymar makes more decisions that get a very low or average rating in tense situations, at the cost of fewer decisions with a high rating. In contrast, Neymar's executions do not suffer under pressure as they improve for all action types.

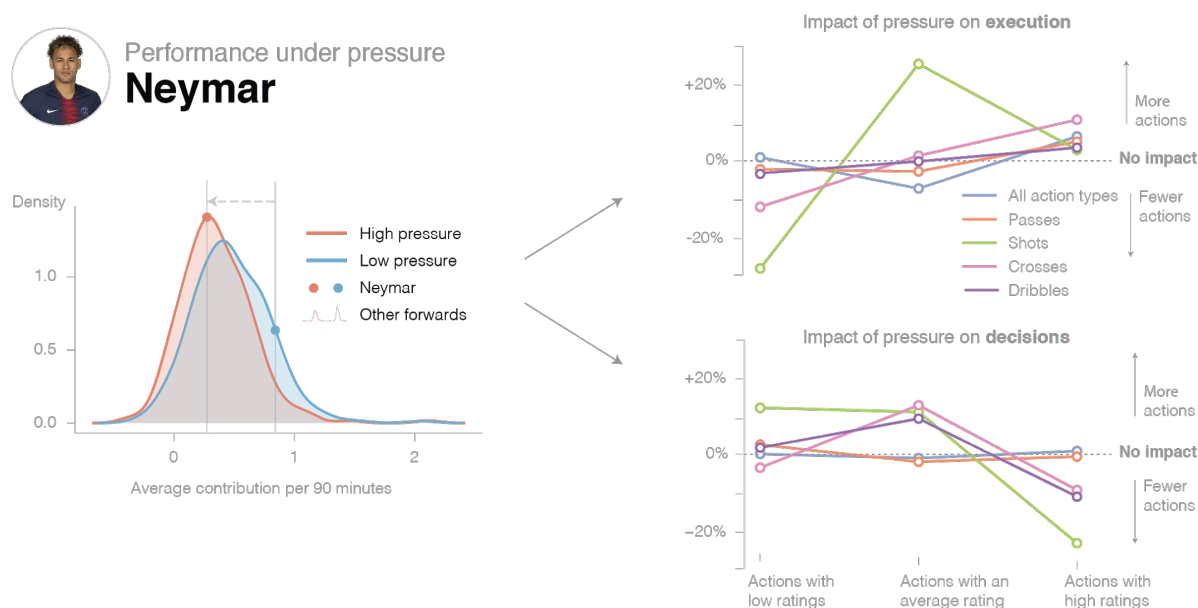


Figure 9. Under pressure, Neymar evolves from a player with a very high average contribution per 90 minutes to an average player. This decrease in performance can mainly be attributed to poor decision making in high-pressure situations.

6.2. MLS stars of the future

Recently, European clubs have been increasingly scouting MLS, as exemplified by Bayern Munich's signing of Alphonso Davies and RB Leipzig's signing of Tyler Adams. Atlanta United's attacking

midfielder Ezequiel Barco is another MLS talent that could soon move overseas. These players represent the top three players aged under 20 in terms of average contribution per 90 minutes in the 2018 MLS season.

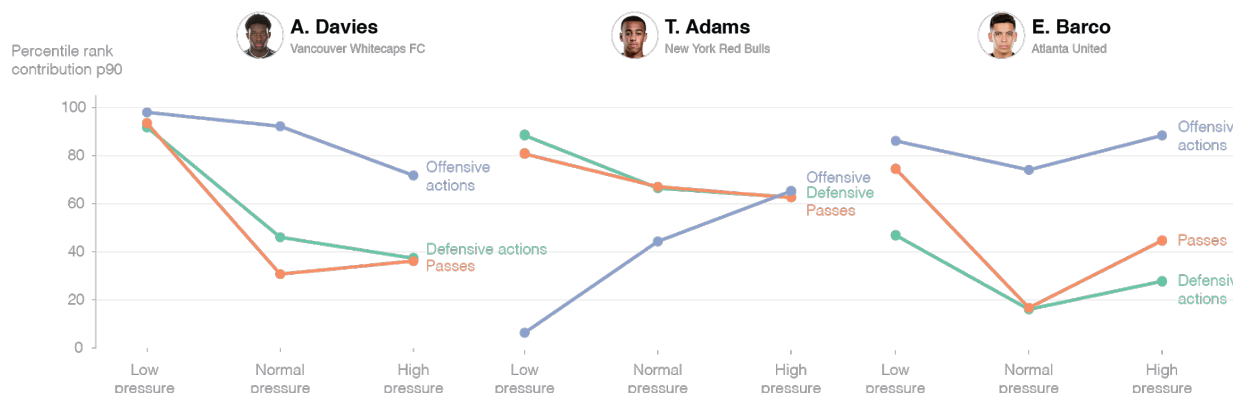


Figure 10. The percentile rank contribution ratings for MLS talents Alphonso Davies, Tyler Adams and Ezequiel Barco under different pressure levels.

Figure 10 shows the percentile rank for the contribution ratings from defensive actions (i.e., interceptions, tackles, clearances and fouls), offensive actions (i.e., crosses, shots, dribbles and take-ons) and passes for all three players under different pressure levels for the 2016, 2017, and 2018 seasons. Since the distribution of the contribution per 90 minutes differs per type of action, we show the percentile ranks for each group of action types. Davies (1271 high-pressure minutes) seems not to cope well with pressure: he is among the best players in the league in low-pressure situations whereas he is a below-average player under high pressure. His contribution per 90 minutes from offensive actions, defensive actions and passes drops as pressure increases. In contrast, Adams (1804 high-pressure minutes) and Barco (760 high-pressure minutes) have more of a mixed reaction to pressure: some actions improve and some decline.

6.3. Liverpool recruitment

Our analysis of Liverpool's signings since the start of the 2017/2018 season reveals that the Reds have a keen eye for buying players who shine under mental pressure. For Alisson (AS Roma), Alex Oxlade-Chamberlain (Arsenal), Andrew Robertson (Hull City), Fabinho (Monaco), Naby Keïta (RB Leipzig), Virgil van Dijk (Southampton) and Xherdan Shaqiri (Stoke City), Figure 11 shows each player's average contribution per 90 minutes in low- and high-pressure situations at their previous clubs. Except for Keïta, each of these signings obtained higher contribution ratings in high-pressure than in low-pressure situations. We omitted Mohamed Salah from our analysis since he played too few minutes under high mental pressure at his former club AS Roma.

Recruitment Liverpool

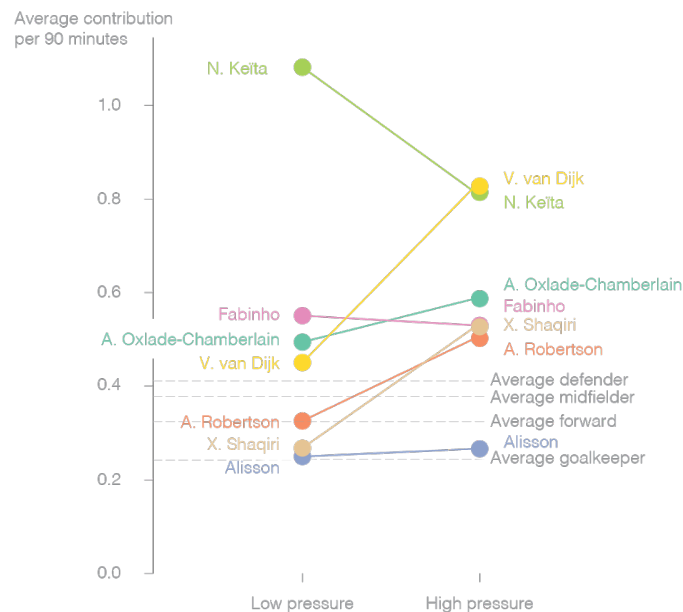


Figure 11. Average contribution per 90 minutes of Liverpool's recent signings in low- and high-pressure situations at their former clubs. For comparison, we include the average contributions of an average player (by position) irrespective of the mental pressure situation. Since the 2017/2018 season, Liverpool has almost consistently signed players who perform better in high-pressure situations. Naby Keita is the only exception.

6.4. Fatigue

Mental pressure, like fatigue, often increases during the game. Therefore, we run an analysis to verify that our findings go beyond detecting performance changes due to fatigue. To counter this argument, Figure 12 compares each player's performance under mental pressure (measured as the player's average contribution per 90 minutes in high-pressure situations) with his performance under fatigue (measured as the player's average contribution per 90 minutes in the last quarter of each game, excluding the high-pressure situations). Based on the weak Pearson correlation of 0.136, we conclude that our findings reflect something different from the fatigue that sets in towards the end of a game.

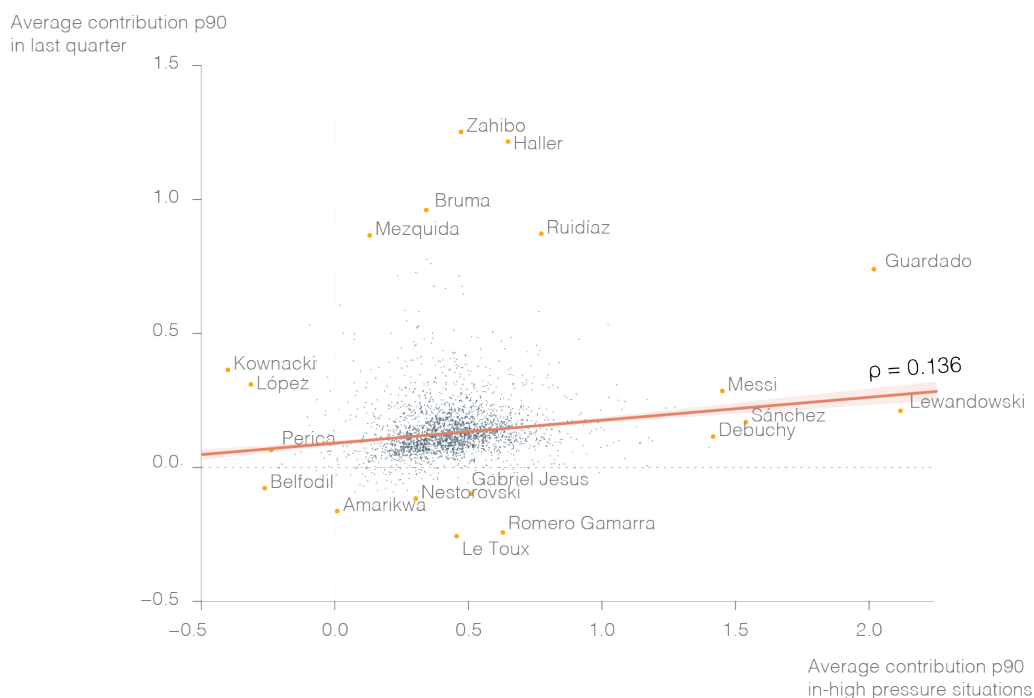


Figure 12. A comparison between each player's performance in high-pressure situations and performance under fatigue (i.e., in the last quarter of a game, excluding the high-pressure situations). Performance is measured as the average contribution per 90 minutes. The low Pearson correlation of 0.136 indicates that our pressure metric does not measure fatigue.

7. Related Work

The effects of pressure on performance have been studied in other contexts and sports. One approach is to define high-pressure situations by hand, as is done in the NBA (e.g., games within five points with five minutes left) or baseball (e.g., runners in scoring position), and then compute performance metrics (both traditional and advanced) for players in such situations. Some attempts in basketball⁶ and baseball^{7,8} consider in-game win probability models to identify high-pressure (or “clutch”) situations. However, they only consider a limited number of actions (e.g., ignore passes, blocked shots, fouls). Our approach differs in several important ways. First, our pressure model considers both pre-game and in-game factors. Second, it considers a broader set of actions, including both offensive and defensive ones. Third, it rates three different aspects of each action. Beyond that, performance under mental pressure has been extensively studied in the cognitive sciences' literature regarding its mechanisms and moderators [11]. This research is mostly experimental and induces pressure through artificial manipulation, distraction and self-focus [10].

⁶ <http://www.inpredictable.com/2014/03/measuring-clutch-play-in-nba.html>

⁷ <https://www.baseballprospectus.com/news/article/38398/prospectus-feature-revised-look-clutch-hitting-part-1/>

⁸ <https://www.baseballprospectus.com/news/article/38519/prospectus-feature-revised-look-clutch-hitting-part-2/>

The component parts of our pressure model have also been considered to varying degrees. Pre-game pressure has been considered in the context of elimination games in NBA playoffs to research how the threat of severe losses affects performance [16]. This research defines a high-pressure game as one where the team faces elimination in case of a loss. They do not take multiple features into account, as we do. In-game win-probability models have been extensively studied in basketball [9], American football [17], and baseball [19]. In contrast, they have received less attention in soccer [3], probably because its low-scoring nature makes it more challenging to analyze. Additionally, there is a long line of work focused on designing advanced metrics for valuing actions and decisions both in soccer [4, 7, 14] and other sports [2, 6, 8, 15].

8. Conclusions

This paper has taken a step towards trying to objectively understand how high-mental pressure situations affect the performances of soccer players. Given that soccer players are often confronted with such situations, more insight into the link between pressure and performance could help in numerous ways. We illustrated concrete use cases about how it could inform acquiring players, coaching individual players, making tactical decisions, and deciding on lineups or substitutions. Some of our findings may have direct practical implications. For example, our analyses indicate that almost all teams could benefit by adjusting how they treat throw-ins. Furthermore, we made several interesting observations, such as that Liverpool, whether by design or not, has recently targeted players who excel under pressure and that Neymar's performances declined when under pressure.

Acknowledgements

We thank our panel of 19 soccer experts for ranking soccer games in terms of pre-game pressure. Furthermore, we thank Max Adema, Tom Decroos, William Spearman, Niels van den Hoek, Yoram Volp, and Albrecht Zimmermann for their comments and suggestions that helped improve this paper. Pieter Robberechts is supported by the EU Interreg VA project Nano4Sports. Jesse Davis is partially supported by the EU Interreg VA project Nano4Sports and the KU Leuven Research Fund (C14/17/07, C32/17/036).

References

- [1] Roy F. Baumeister. "Choking under Pressure: Self-Consciousness and Paradoxical Effects of Incentives on Skillful Performance." In: *Journal of Personality and Social Psychology* 46.3 (1984), pp. 610–620.
- [2] Benjamin S. Baumer, Shane T. Jensen, and Gregory J. Matthews. "openWAR: An Open Source System for Evaluating Overall Player Performance in Major League Baseball." In: *Journal of Quantitative Analysis in Sports* 11.2 (2015), pp. 69–84.
- [3] Jay Boice. "How Our 2018 World Cup Predictions Work." *FiveThirtyEight* (June 2018). <https://fivethirtyeight.com/features/how-our-2018-world-cup-predictions-work/>.
- [4] Lotte Bransen and Jan Van Haaren. "Measuring Football Players' On-the-Ball Contributions From Passes During Games." In: arXiv:1810.02252 [cs, stat] (September 2018).
- [5] Shael Brown. "A PageRank Model for Player Performance Assessment in Basketball, Soccer and Hockey." In: *Proceedings of the 11th MIT Sloan Sports Analytics Conference*. Boston, 2017, p. 22.



- [6] Dan Cervone, Alexander D'Amour, Luke Bornn, and Kirk Goldsberry. "Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data." In: Proceedings of the 8th MIT Sloan Sports Analytics Conference. 2014, p. 9.
- [7] Tom Decroos, Lotte Bransen, Jan Van Haaren, and Jesse Davis. "Actions Speak Louder Than Goals: Valuing Player Actions in Soccer." In: arXiv:1802.07127 [stat] (February 2018).
- [8] Football Outsiders. Methods To Our Madness. <https://footballoutsiders.com/info/methods>.
- [9] Sujoy Ganguly and Nathan Frank. "The Problem with Win Probability." In: Proceedings of the 12th MIT Sloan Sports Analytics Conference. Boston, 2018.
- [10] Daniel F. Gucciardi and James A. Dimmock. "Choking under Pressure in Sensorimotor Skills: Conscious Processing or Depleted Attentional Resources?" In: Psychology of Sport and Exercise 9.1 (2008), pp. 45–59.
- [11] Denise M. Hill, Sheldon Hanton, Nic Matthews, and Scott Fleming. "Choking in Sport: A Review." In: International Review of Sport and Exercise Psychology 3.1 (2010), pp. 24–39.
- [12] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M. Blei. "Automatic Differentiation Variational Inference." In: arXiv:1603.00788 [cs, stat] (March 2016).
- [13] Jiří Lahvička. "Using Monte Carlo simulation to calculate match importance: The case of English Premier League." In: Journal of Sports Economics 16.4 (2015), pp. 390–409.
- [14] Patrick Lucey, Alina Bialkowski, Mathew Monfort, Peter Carr, and Iain Matthews. "Quality vs Quantity: Improved Shot Prediction in Soccer using Strategic Features from Spatiotemporal Data." In: Proceedings of the 8th MIT Sloan Sports Analytics Conference. Boston, 2014.
- [15] Patrick McFarlane. "Evaluating NBA End-of-Game Decision-Making." In: Journal of Sports Analytics. Preprint (2018), pp. 1–6.
- [16] Elia Morgulev and Yair Galily. "Choking or Delivering Under Pressure? The Case of Elimination Games in NBA Playoffs." In: Frontiers in Psychology 9 (June 2018).
- [17] Konstantinos Pelechrinis. "iWinRNFL: A Simple, Interpretable & Well-Calibrated In-Game Win Probability Model for NFL." In: arXiv:1704.00197 [stat] (April 2017).
- [18] Andrew S. Phelps, David M. Naeger, Jesse L. Courtier, Jack W. Lambert, Peter A. Marcovici, Javier E. Villanueva-Meyer, and John D. MacKenzie. "Pairwise Comparison Versus Likert Scale for Biomedical Image Assessment." In: American Journal of Roentgenology 204.1 (2015), pp. 8–14.
- [19] Alan Schwarz and Peter Gammons. The Numbers Game: Baseball's Lifelong Fascination with Statistics. Reprint edition. Thomas Dunne Books, October 2013.
- [20] William Spearman. "Beyond Expected Goals." In: Proceedings of the 12th MIT Sloan Sports Analytics Conference. Boston, 2018.
- [21] William Spearman, Austin Basye, Greg Dick, Ryan Hotovy, and Paul Pop. "Physics-Based Modeling of Pass Probabilities in Soccer." In: Proceedings of the 11th MIT Sloan Sports Analytics Conference. Boston, 2017.



Appendix

A.1 Evaluation of the pre-game pressure model

The lack of a ground truth is the major challenge in this task. Therefore, we leveraged the expertise of a panel of 19 soccer experts to learn and evaluate our ranking classifier. From this panel, we obtained 330 pairwise rankings from a randomly selected set of 170 games from the 2016/2017 and 2017/2018 Premier League, Bundesliga and LaLiga seasons. We used these pairwise rankings to fit a Gradient Boosted Ranking Trees model (*max_depth*: 6, *n_estimators*: 50). Additionally, to validate our model, we obtained 483 pairwise ratings for a diverse set of 20 games, including some crucial relegation games, rivalries, games of teams underperforming and end-of-season games where nothing was at stake anymore. An accurate model should come close to mimicking the experts' aggregate (partial) ordering of these matches' pressure levels. However, the experts may not rank each pair in the same way (i.e., there is no consensus ordering), so an accurate model should perform similarly to the inter-expert agreement. Our learned model achieved an agreement of 73.91% with the annotators' rankings, which is close to the inter-expert agreement of 79.79%.

A.2 Specification and evaluation of the in-game win probability model

As mentioned in the main body of the article, we model the future number of goals that the home team ($y_{>t,home}$) and away team ($y_{>t,away}$) will score as independent Binomial distributions:

$$y_{>t, home} | \theta_{t, home} \sim B(T-t, \theta_{t, home}) \text{ and} \\ y_{>t, away} | \theta_{t, away} \sim B(T-t, \theta_{t, away}),$$

where the θ parameters represent the estimated scoring intensity in the t^{th} time frame for the teams playing at home and away, respectively. These scoring intensities are estimated from the current game state features $x_{t, home}$ and $x_{t, away}$, which are described in the main body of the article. Since the importance of these game state features varies over time in a non-linear way, we model the scoring intensities using a temporal stochastic process. Specifically, we use the following specifications and priors to model the scoring intensities:

$$\begin{aligned} \text{logit } \theta_{t, home} &= \alpha_{t, home} + \beta + H_a & \text{with } \alpha_{t, home} &\sim N(\alpha_{t-1}, \sigma^2) \\ \text{logit } \theta_{t, away} &= \alpha_{t, away} + \beta & \beta &\sim N(0, 10) \\ & & H_a &\sim N(0, 10), \end{aligned}$$

where H_a models the home advantage.

We estimate and validate the corresponding model on event data from the 2014/2015 to 2017/2018 seasons of the English Premier League, Spanish LaLiga and German Bundesliga seasons (4232 games in total). Therefore, we use a 70-30% train-test split that respects the temporal ordering of games. To assess the quality of our model, we calculate for all games where our model predicts a win, draw or loss probability of $x\%$ the fraction of games that actually ended up in that

outcome. The probability calibration curves in Figure A.1 show how much our predicted probabilities deviate from the actual probabilities. Our model stays close to the 0% line, which is the ideal (i.e., best possible) performance. The estimates are typically within 2.5% of this line and have a maximum deviation of 9.61%.

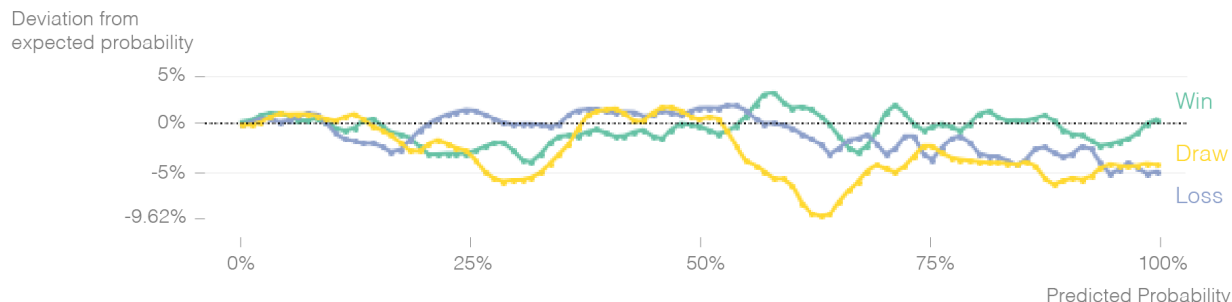


Figure A.1. Probability calibration curves for the in-game win probability model. The predicted probabilities match very well with the actual outcome, proving that our model is well calibrated.

A.3 Evaluation of the action success model

To predict the probability that a given action will be successful, we train a binary Gradient Boosted Trees classifier on historical match data from the 2014/2015 and 2015/2016 seasons of the English, Spanish, German, Italian, French, Dutch and Belgian first divisions. After removing games for which relevant data is missing (such as the action type), we obtain a training set of 2430 games and 2.96 million actions in total. We validated the model on the remaining 2016/2017 seasons of the same leagues, comprising 2404 games and 2.92 million actions.

Since the definition of success differs per action type, we train a separate model for each of them. Table A.1 lists the Area Under the Precision Recall Curve (AUC-PR) for these models and compares it to the base rate of success of the corresponding action type.

Table A.1. The Area Under the Precision Recall Curve (AUC-PR) for each of the Gradient Boosted Trees classifiers that predict whether an action of a given type will be successful. The base rate of success is provided as a baseline.

Action type	Base rate AUC-PR on validation set (equals the base rate)	Our model's AUC-PR on validation set
Pass	0.8265	0.9561
Cross	0.3144	0.5397
Throw-in	0.8821	0.9677
Corner (crossed)	0.4582	0.6550
Corner (short)	0.8514	0.9829

Freekick (crossed)	0.5612	0.7429
Freekick (short)	0.8761	0.9783
Freekick (shot)	0.0668	0.1011
Take on	0.6991	0.8357
Tackle	0.4575	0.7267
Shot	0.1094	0.3689
Penalty	0.7480	0.7480
Dribble	0.9916	0.9916
Goal kick	0.5665	0.8310

Besides a high accuracy, we need well-calibrated probabilities in order to use these success probability estimates in our decision and execution metrics. Therefore, we analyze the calibration curve for all action types combined (Figure A.2). Since the deviation from the true success probabilities stays within 3%, we can conclude that our model is well calibrated.

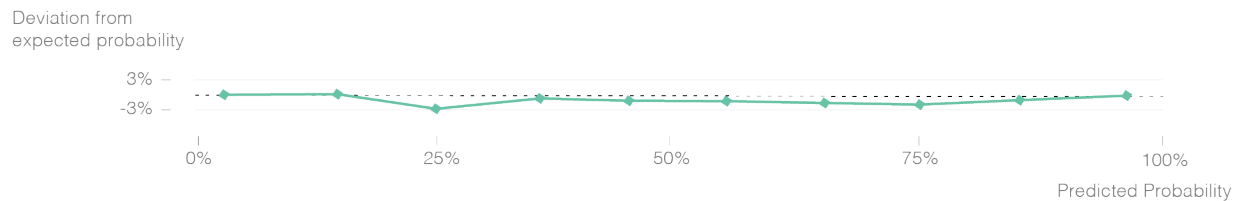


Figure A.2. Probability calibration curve for the action success model. The predicted success probabilities match very well with the actual action outcomes, proving that our model is well calibrated.

A.4 Evaluation of estimating the next action's contribution rating

Two Gradient Boosted Trees models underlie our estimations of the next action's contribution rating. The first model estimates the probability $P(gi+1+ | si)$ of scoring a goal in the next game state $si+1$ given the current game state si ; the second model estimates the probability $P(gi+1- | si)$ of conceding a goal in the next game state $si+1$ given the current game state si . These probabilities are combined as follows:

$$ECR | si = [P(gi+1+ | si) - P(gi+ | si-1)] - [P(gi+1- | si) - P(gi- | si-1)].$$

We train and evaluate both models on the same datasets as used in Appendix A.3. Table A.2 shows the Mean Squared Error (MSE) for both models on the validation set. As a baseline we use the

average next game state value of all actions starting in the same location.

Table A.2. The Mean Squared Errors (MSE) for the models that estimate the probability of scoring or conceding a goal in the next game state. The average next game state value of all actions that start in the same location serves as a baseline.

Model	Baseline MSE	Our model's MSE
$P(g_{i+1+} \mid s_i)$	0.0023	0.0019
$P(g_{i+1-} \mid s_i)$	4.9139e-06	4.0413e-06