BARÇA INNOVATION HUB

FOOTBALL ANALYTICS 2021

The role of context in transferring analytics to the pitch

Editors
Angel Ric
Sam Robertson
David Sumpter

Section leaders
Paul Bradley
Lotte Bransen
Isaac Guerrero
Matthias Kempe
Carlos Lago
Maurici A. López-Felip
Rafel Pol
Laurie Shaw
Hans Thies

EDITORS

Angel Ric
Sam Robertson
David Sumpter

DESIGNER AND PUBLISHER

Barça Innovation Hub

SECTION LEADERS

Paul Bradley
Lotte Bransen
Isaac Guerrero
Matthias Kempe
Carlos Lago
Maurici A. López-Felip
Rafel Pol
Laurie Shaw
Hans Thies

INTERNATIONAL CONTRIBUTORS

Jack D. Ade
Natàlia Balagué
Pascal Bauer
Jesse Davis
Xavier Damunt
Tom Decroos
James A. Dixon
Bahaeddine El Fakir
Mark Evans
Nicolas Evans
Jordi Fernández
Floris Goes
Miguel Ángel Gómez
Antonio Gómez-Díaz
Jan Van Haaren
Henry S. Harrison
Mat Herold
Wonwoo Ju
Andy Laws
Matt N. Petersen
Richard Pollard
Marc Quintana
Pieter Robberechts
Paco Seirul·lo

AFFILIATIONS

Preface
Raúl Peláez
FC Barcelona

Introduction
Angel Ric
Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), Lleida, Spain
FC Barcelona

Sam Robertson
Victoria University, Australia

David Sumpter
Uppsala University
Hammarby IF

Chapter 1
Rafel Pol
Real Federación Española de Fútbol (RFEF)
Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), University of Lleida, Spain

Natàlia Balagué
Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), University of Barcelona, Spain

Chapter 2
Hans Thies
FIFA

Bahaeddine El Fakir
FIFA

Nicolas Evans
FIFA

Chapter 3
Carlos Lago
University of Vigo, Spain

Miguel Ángel Gómez
Technical University of Madrid, Spain

Richard Pollard
California Polytechmic State University, United States

Chapter 4
Lotte Bransen
SciSports

Pieter Robberechts
KU Leuven

Jesse Davis
KU Leuven

Tom Decroos
KU Leuven

Jan Van Haaren
KU Leuven
SciSports

Chapter 5
Laurie Shaw
Department of Statistics, Harvard University

Chapter 6
Matthias Kempe
Department of Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), Groningen, The Netherlands.

Floris Goes
Department of Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), Groningen, The Netherlands.

Mat Herold
Institute of Sports and Preventive Medicine, Saarland University, Saarbrücken, Germany

Pascal Bauer
German Football Federation (DFB)

Chapter 7
Isaac Guerrero
FC Barcelona

Xavier Damunt
FC Barcelona

Marc Quintana
FC Barcelona

Jordi Fernández
FC Barcelona

Paco Seirul·lo
FC Barcelona

Chapter 8
Maurici A. López-Felip
Center for the Ecological Study of Perception and Action, Department of Psychological Sciences, University of Connecticut, United States
FC Barcelona

Matt N. Petersen
Center for the Ecological Study of Perception and Action, Department of Psychological Sciences, University of Connecticut, United States

Henry S. Harrison
Kognia Sports Intelligence

James A. Dixon
Center for the Ecological Study of Perception and Action, Department of Psychological Sciences, University of Connecticut, United States

Chapter 9
Paul Bradley
Football Science Consultant

Wonwoo Ju
Researcher in Football Science at LJMU

Jack D Ade
U23's Fitness Coach at Liverpool Football Club

Andy Laws
Computer Scientist at LJMU

Antonio Gómez-Díaz
FC Barcelona

Mark Evans
Computer and Data Scientist

BARÇA INNOVATION HUB

Football Analytics 2021

The role of context in transferring analytics to the pitch

Overview

Biographies

ANGEL RIC, Ph.D.

Angel Ric has a doctorate in physical activity and sport sciences from the University of Lleida (UdL), a Master's degree in high performance in team sport (Mastercede, Byomedic System), and is a national football coach. Currently, he is football professor at the National Institute of Physical Education of Catalonia (INEFC). Since 2014 he has been a member of the Complex Systems and Sport Research Group. His research focuses on studying football as a complex phenomenon and analysing the dynamics of coordination between players in competitive environments. He also belongs to the sports science department of FC Barcelona in the area of technology, analysis and innovation, and collaborates on different projects for the generation and dissemination of knowledge with the Barça Innovation Hub.

SAM ROBERTSON, Ph.D.

Sam Robertson is a professor of sports analytics at Victoria University in Melbourne, Australia where he also leads research, partnership and commercial activity across sport. Both his research and practice predominantly focus on the application of different analytical approaches to various sports performance problems. Both Sam and his team of analysts, researchers and students undertake projects and consultancy with professional franchises in sports such as professional football, basketball, baseball and Australian football. Sam also works with a number of governing bodies including FIFA, Tennis Australia and the Australian Football League (AFL) on topics relating to developing new match insights, driving performance and commercial innovation as well as the application and validation of performance technologies.

DAVID SUMPTER, Ph.D.

Professor David Sumpter is author of *Collective Animal Behaviour* (2010), *Soccermatics* (2016), *Outnumbered* (2018) and *The Ten Equations that Rule the World* (due 2020). He has published over 100 scientific articles in leading journals on everything from the inner workings of fish schools and ant colonies, through social psychology and segregation in society, to machine learning and artificial intelligence. He has consulted for leading football clubs, in betting and works actively with outreach to schools, industry and the social sector. His talks at Google, TedX, the Oxford Mathematics Public Lecture and The Royal Institution are available online.

RAFEL POL

Rafel Pol has a degree in physical activity and sport sciences from the University of Barcelona (UB), a Master's degree in high performance sports (UB) and a Master's in prevention and physical-sports readaptation of football injuries from the University of Castilla la Mancha. He is currently strength & conditioning coach of the Spanish national football team. In the past, he has held the same role in AS Roma, RC Celta de Vigo and FC Barcelona. He is a member of the Complex Systems and Sport Research Group and his research focuses on the study of sport as a complex phenomenon.

HANS THIES, Ph.D.

Hans is an IT programme manager at FIFA's Member Associations division. He held several positions in consulting, IT project management and IT research at companies such as Porsche and SAP before joining FIFA in 2014. He has a Master's degree in engineering from the Karlsruhe Institute of Technology (KIT) and a PhD in Business Innovation from the University of St. Gallen (HSG). At FIFA, he is responsible for a digital football development programme in the Member Association division. The programme he leads aims to streamline, standardise, and digitalise existing processes and the related data in the administration of the sport; harmonise and connect IT systems in football; facilitate the exchange and enrichment of data in the industry; and enable new business models and revenue generation through IT systems and data.

CARLOS LAGO, Ph.D.

Carlos Lago-Peñas (@ Clagopuvigo) is a full professor of football and performance analysis in sport at the University of Vigo, Spain. Both his research and practice are focused on performance analysis in team sports. He is consultant in the topic of performance analysis for several high-level clubs. He is associate editor for the *International Journal of Performance analysis in Sport*. He currently collaborates with Barça Innovation Hub and LaLiga.

LOTTE BRANSEN

Lotte Bransen is a lead data scientist at SciSports, where she leads the data analytics team that develops analytical tools to derive actionable insights from football data. An avid football player herself, Lotte primarily works on developing machine learning models to measure the impact of football players' in-game actions and decisions on the courses and outcomes of matches. Prior to SciSports, Lotte obtained a Master of Science degree in Econometrics & Management Science from Erasmus University Rotterdam and a Bachelor of Science degree in Mathematics from Utrecht University.

LAURIE SHAW, Ph.D.

Dr. Laurie Shaw is a lecturer and research scientist in the department of statistics at Harvard University, and a Fellow of the Harvard Data Science Initiative. He holds a PhD in computational astrophysics and has previously worked in quantitative finance and for the British Government. His research currently focuses on sports analytics, applying statistical and machine learning techniques to extract information from data and creating tools to help professional teams make use of this information.

MATTHIAS KEMPE, Ph.D.

Dr. Matthias Kempe is an assistant professor of data science in sports at the University of Groningen. He received his PhD in Sport Science at the German Sport University Cologne form the Faculty of Exercise Training and Sport Informatics. His research interests include performance optimisation and decision making in team sports as well as sports analytics. This work has resulted in publications in journals such as *Big Data, Journal of Sport Science, European Journal of Sport Science*, and *Experimental Aging Research*.

ISAAC GUERRERO

Isaac Guerrero is the deputy director & head of coaching at the FC Barcelona Methodology Area. Guerrero has been the technical director of the FC Barcelona Football School since 2010 and is currently also a member of the Knowledge Area as well as being part of the Barça Innovation Hub, all in FC Barcelona. During the last 5 seasons, he has also been the coordinator of the Barça Coach Development Programme and during the last two seasons has been director and professor of the FC Barcelona Professional Football Coach Master. Guerrero has a degree as a physical education teacher (extraordinary award as number 1, class of 2003) and a science degree in physical activity and sport, both obtained from the University of Barcelona (UB). He is also a Spanish Football Federation Level 3 and UEFA Pro qualified coach. He is co-author of the book *The Decisional Emergence Influenced by the Emotions. Working a Proposal for Improving Decision Making in Youth Football* and the upcoming book *The Creation of a Non-Verbal Communication Code for Football* as well as speaker at seminars and conferences.

MAURICI A. LÓPEZ-FELIP, Ph.D.

Maurici A. López-Felip is a behavioural scientist who specialised in ecological physics and earned his PhD at the University of Connecticut (UConn). His research focuses on formalising physical and systemic nature to quantify the principles underlying the game. Maurici also acts as a research consultant for different football clubs and sport science institutions, including FC Barcelona, AIK Football, the Simulation, Training, Analytics and Rehabilitation (STAR) Heel Performance Laboratory at the University of North Carolina, and the Center for the Ecological Study of Perception and Action at UConn. He also organises an international graduate programme on science and technology in football, and recently started his own business venture leveraging coaching and scientific insights with technological-based solutions for practitioners.

PAUL BRADLEY, Ph.D.

Dr Paul Bradley specialises in integrative football solutions (i.e. fusing tracking and tactical data together). He conducts translational consultancy work in an elite football setting that bridges the gap between cutting-edge research and professional practice (i.e. not just research for research's sake but R&D that adds value to the applied setting). In addition to working as a consultant to elite clubs he is also very active in R&D and has published >65 peer-reviewed papers in the area of science and football that have acquired >4000 citations. He is also an associate professor (reader) at LJMU.

Preface

— Raúl Peláez[1]

1: FC Barcelona

INTRODUCTION

Football is a very complex sport to be understood and explained in a simple and linear way. Many factors provoke countless responses that are nested and create constant adaptations. Football is a complex adaptive system at different scales, in a constant balance between order and disorder. Therefore, football is mainly about players and the interaction between them and their environment.

Coaches generate stories around an idea of the game, a competition, an opponent, a match; and players have the ability to believe these stories and cooperate with each other to achieve a goal. The success of a team depends a lot on this story and how credible it is to the players. The emotions aroused by this interaction among team members (including the coaching staff) means that these behaviours derived from the style of play can be stabilised or, on the contrary, diluted.

Why should a player or a coach believe more in data or an algorithm than in an expert opinion that explains a specific situation of the game? Why are a couple of subjective phrases from a coach more successful than a table of objective data from a scientist?

We must know that football is infinitely more complex than a set of computations and instructions, of ordered and finite rules that try to give an answer to a problem or solve an equation. It will be difficult, or almost impossible, to control the game from mathematics, but the closer we are to understanding the context and contextualising the data, the closer we will be to the game and the players.

On the other hand, we must recognise that people today know less than we think they do. Moreover, we do not think as individuals but rather in groups, as we treat the knowledge of others as if it were our own. We humans are very limited thinkers on an individual level, but very powerful on a collective level.

I would like to invite data scientists who want to devote themselves to the world of football to think about their role. The document that the reader is about to discover scientifically justifies the importance of contextualising data. But this is only the beginning. The data cannot be thrown directly onto an analyst's or a coach's desk; that would be the quickest way for it to end up in the bin if delivered on paper, or in the computer trashcan if delivered digitally. Reports must also be accompanied by a story that makes them credible. That is to say, in addition to computing with the maximum rigour, and paying attention to the context, it will also be appropriate to analyse the results in a group and to collaborate with the sport's practitioners to try to reach a conclusion together that provides added value. In this way we can generate a short story that is both intelligible and agile.

Data has arrived in football, just as it has arrived in other areas of our lives in recent years: invasively and almost always in a contradictory way. This leads to its rejection in some professional environments. Therefore, an effort must be made to make sense of the information and to discern what is important from what is not.

Above all, a data scientist who wants to succeed in the world of sport must have, beyond technical knowledge, highly developed skills in areas such as:

An understanding of the phenomenon being studied, which allows for analyses tailored to the reality of the end-user.

- Collaboration, to be able to think with others, to learn, to optimise their work and to evaluate knowledge.
- Communication, to be able to explain their knowledge and findings, to integrate into multidisciplinary teams and to find the right moments to communicate each specific aspect.
- Creativity, to explore new scenarios, to reinvent themselves, to give answers to questions, and to redirect their theories.
- Critical thinking, to reason and reflect without being influenced by power; to know how to discriminate what is important from what is not; and to make decisions and solve problems in critical or extreme situations.

Enjoy reading this publication without forgetting to know yourself and to think about what you want in life. Always keep in mind the ability to control technology, before technology controls you.

Introduction to football analytics 2021

— Angel Ric[1,2], Sam Robertson[3] & David Sumpter[4,5]

1: Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), Lleida, Spain

2: FC Barcelona

3: Victoria University, Australia

4: Uppsala University

5: Hammarby IF

INTRODUCTION

The incorporation of new technology into sport has resulted in an increase in both the volume of data and the variety of its types. This has led to a need for completely new areas of expertise (such as that provided by data analysts, computer scientists and mathematicians) in organisational departments, as well as adaptation by those already holding established roles, such as sport scientists, strength and conditioning coaches and performance analysts. The link between sports and computer sciences is now well established, and few top-level professional structures resist its integration. However, despite the interdisciplinarity of many staff teams, the knowledge and linguistic barriers between different scientific fields mean that interaction between professionals can often be difficult. One way to help address this problem is for experiential and scientific knowledge, historically based on reductionism, to move towards an understanding of sport as a complex phenomenon.

The complexity of football lies mainly in the number of interactions between teammates and their environment, in particular their opponents, but is also found in the context of the game as a whole. Traditionally, performance has been evaluated by isolating system components including the physical and physiological; the technical (derived from the players' on-ball events); or tactical aspects (the relationship between teammates and opponents). These "performance components" cannot be seen at the same level but instead in a nested way, because the inherent processes within them emerge on different timescales (figure 1). It is in a competitive environment, evolving over several years, where cooperative processes are situated. The team cooperates at the timescale of weeks and months as well as more immediately at the timescale of the match's 90-plus minutes to achieve the main objective of winning. Interpersonal synergies between players are then formed and dissolved on a scale of seconds and tens of seconds on the pitch, allowing them to score a goal or avoid conceding (Ric et al. 2016). These team patterns then make the context for the individual actions that emerge in a smaller timescale (seconds or milliseconds). Therefore, context must be understood as the environment for an organism-environment relationship at all scales and in performance analysis, from seeing the team as an organism (Duarte et al. 2012), through to the player himself within the team, and even down to the organs or systems within the athlete (Pol et al. 2020).

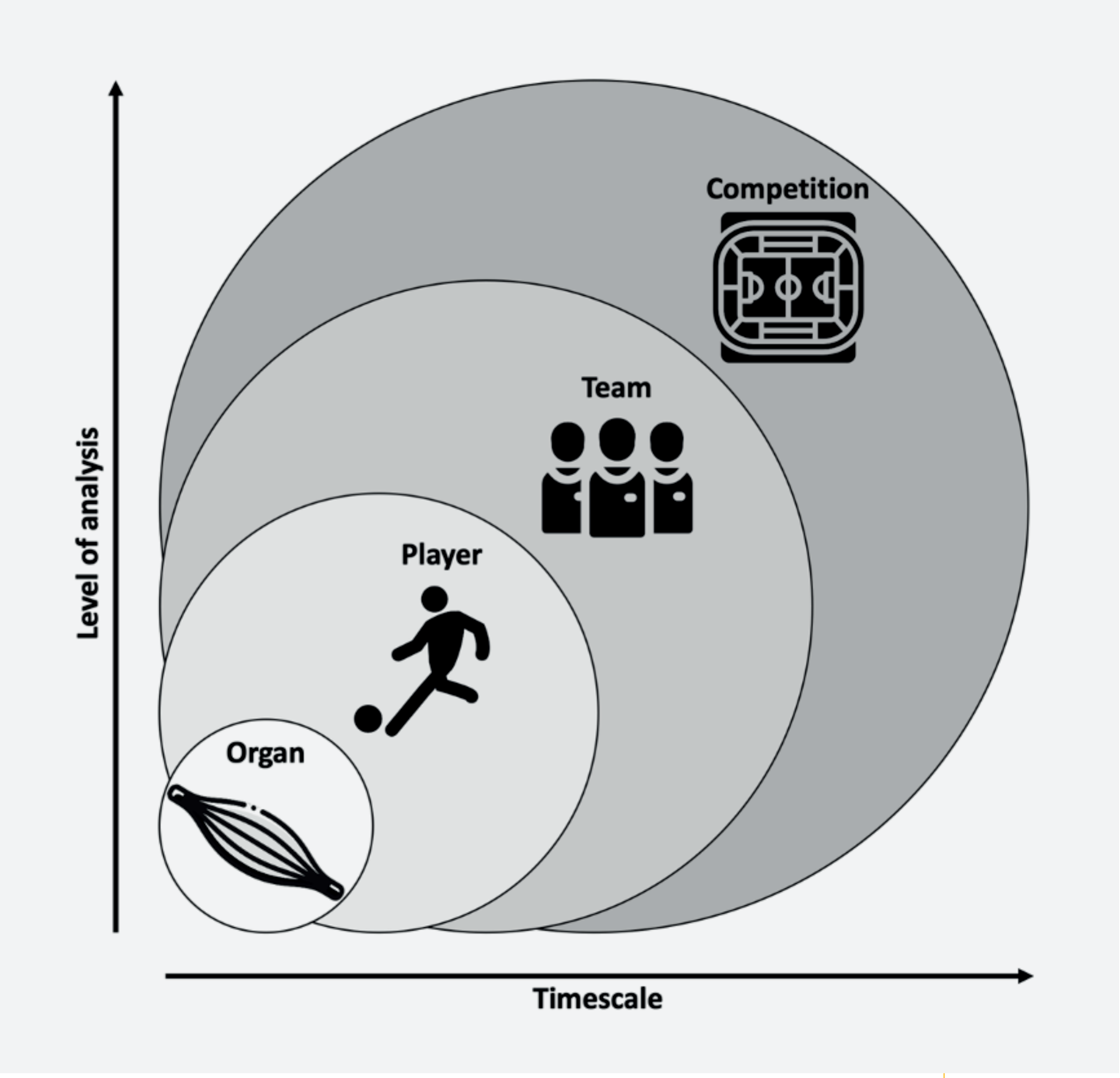


Figure 1. Nested levels and scales of performance analysis.

For these levels of performance analysis, tracking data (team), event data (player) and biometric data (organ) is collected systematically. The integration of many of these data sets is essential and has been recommended within clubs or institutions for proper data management (Linke et al. 2018). FIFA, in its aim to assist associate members or other stakeholders, is developing standards for the unification of these data sources. Related to this, the assignment of unique player IDs aims to avoid duplication and incompatibility between different data sources and competitions, thus providing simplicity for joint data analysis at all levels of performance. At a higher level, where performance is affected by competition (e.g. the league), the effects of situational or contextual variables on the behaviour of players and teams should be considered (Gómez et al. 2013). In fact, during FIFA competitions, performance can be affected by macro-considerations such as the location and climate of where the competition takes place. However, variables that influence at shorter time scales (e.g. match location, half, score, team formation, etc.) have also been extensively studied. Traditionally, these variables have been represented as frequencies of individual actions or aggregations of collective behaviour derived from event and tracking data respectively. However, this type of analysis makes it difficult to consider the context in the description of the results. Consequently, recent statistical models have emphasised estimating the value of those actions (Decroos et al. 2019), as well as the behaviour of the team's performance in relation to the opponent's disposition and the ball (Fernández & Bornn, 2020). The value of these behaviours (individual or collective) derives from the context in which they take place. For example, the value of actions-decisions with or without a ball has been quantified according to variables such as location, distance to the goal or the angle, that vary in tens of second. On the other hand, the effect of the style and system of play, or the number of players on each line of pressure and defensive organisation, implies contextual changes on a larger scale of seconds or tens of seconds. Furthermore, changes in situational variables, already noted above, play an important role in the effects they can have on mental or emotional pressure (Bransen et al. 2019).

It is common for media and fans to consider players in a position within a system of play or team formation. However, within the same position(s), different roles exist. Because of the dynamism of the game, these positions are not static and depend on specific contexts that can be identified within the game. Among them, moments in which a team has the ball and others in which it tries to recover it can be differentiated. Similarly, playing strategies such as pressure or withdrawal to recover the ball or more elaborate direct play when possessing the ball can also be identified (Castellano & Pic, 2019).

Furthermore, the ball location between the different confrontation lines (Castellano et al. 2007) and the depth of these lines will also determine more specific contexts of interaction. These structures can be practised and learnt over the time scale of months, but can change in the blink of an eye.

Identifying the most common formations during these game contests, and the changes derived from the transitions between them, are some of the challenges that some researchers are already beginning to address. The stability of these formations, and also derived from the values of width and depth of the team, dispersion of the players respect to their team centroid and the depth of the rear line, characterise its organisation. In turn, this definition makes possible to characterise defensive disruption as a function of the variation from the previous set of variables, allowing players' performance to be classified according to the impact that their passes have on the opponent's defensive (dis) organisation (Goes et al. 2019). Most football teams are organised according to the location of the goals, because of two main objectives: to score a goal and protect their own goal. However, the location of the ball is essential to define the context of the game and, in some cases, it is the element that determines the organisation of the players, regardless of the team that is in possession of the ball. So much so that FC Barcelona has been clearly differentiated from other teams (Gyarmati et al. 2014) because of how the team makes use of it through passing, as an action by which the players relate and interact. Although the Barça style of play has been characterised by high passing percentages, this is not an end in itself, but a means through which the players organise themselves around the ball.

It is precisely this organisation that defines one of the most common drills during the club's team training sessions, the "rondo". Despite the extensive research on the use of small-sided games (Clemente et al. 2020), the effects from rondos or other common training drills such as positional games (Casamichana Gómez et al. 2018) on player behaviour or the dynamics of the coordination between them are still unknown. Little research has addressed the need to identify values, parameters or patterns of play that allow the design of training scenarios that are representative of the game. There are already some publications which, through the identification of physical demands (e.g. distance covered) according to the game context, help to design drills that fit with the identified demands (Bradley & Ade, 2018). However, making training drills that meet not only physical but also behavioural demands is still a challenge for both data and sports scientists. This will be best achieved through collaborative efforts.

In summary, one of the challenges for researchers and practitioners is to break down the linguistic barriers between the areas of knowledge or fields of science in order to allow better and more efficient information transmission. The adoption of a transdisciplinary theoretical approach will allow a better understanding of all scales of the emergent processes arising from football environments (e.g. talent identification and development, performance analysis, injury risk assessment, etc.). The identification of essential variables through current analytical techniques (e.g. machine learning) from the inductive approach would allow the modelling of simulations derived from those variables (deductive approach) to capture the system's performance/intelligence (Hristovski & Balagué, 2020). The conceptualisation, quantification and modelling, through entropy, could one day elucidate a paradigm shift in sports performance that might allow its unification at all levels of analysis. Under the nested relationship between levels, the context-dependent analysis of performance will be presented in the following chapters by offering to the reader a wide view of the research carried out by some of the most renowned experts from that topic.

REFERENCES

Bradley, P. S., & Ade, J. D. (2018). Are Current Physical Match Performance Metrics in Elite Soccer Fit for Purpose or is the Adoption of an Integrated Approach Needed? *International Journal of Sports Physiology and Performance,* 13(5), 656–664. https://doi.org/10.1123/ijspp.2017-0433

Bransen, L., Robberechts, Pi., Van Haaren, J., & Davis, J. (2019). Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. *Proceedings of the 13th MIT Sloan Sports Analytics Conference*, 1–25. http://www.sloansportsconference.com/wp-content/uploads/2019/02/Choke-or-Shine-Quantifying-Soccer-Players-Abilities-to-Perform-Under-Mental-Pressure.pdf

Casamichana Gómez, D., Gómez Diaz, A., Cos Morera, F., & Martín García, A. (2018). Wildcard Players During Positional Games. *Apunts. Educacion Fisica y Deportes*, 133, 85–97. https://doi.org/10.5672/apunts.2014-0983.es.(2018/3).133.06

Castellano, J., Hernández-Mendo, A., Morales-Sánchez, V., & Anguera, M. T. (2007). Optimising a Probabilistic Model of the Development of Play in Soccer. *Quality and Quantity*, 41(1), 93–104. https://doi.org/10.1007/s11135-005-3148-0

Castellano, J., & Pic, M. (2019). Identification and Preference of Game Styles in LaLiga Associated with Match Outcomes. *International Journal of Environmental Research and Public Health,* 16(24), 5090. https://doi.org/10.3390/ijerph16245090

Clemente, F. M., Afonso, J., Castillo, D., Arcos, A. L., Silva, A. F., & Sarmento, H. (2020). The effects of Small-Sided Soccer Games on Tactical Behavior and Collective Dynamics: A Systematic Review. *Chaos, Solitons and Fractals*. https://doi.org/10.1016/j.chaos.2020.109710

Decroos, T., Van Haaren, J., Bransen, L., & Davis, J. (2019). Actions Speak Louder Than Goals: Valuing Player Actions in Soccer. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/3292500.3330758

Duarte, R., Arajo, D., Correia, V., & Davids, K. (2012). Sports Teams as Superorganisms: Implications of Sociobiological Models of Behaviour for Research and Practice in Team Sports Performance Analysis. *Sports Medicine*, 42(8), 633–642. https://doi.org/10.2165/11632450-000000000-00000

Fernández, J., & Bornn, L. (2020). SoccerMap: A Deep Learning Architecture for Visually-Interpretable Analysis in Soccer. *Proceedings of the 2020 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 1–16. https://bitbucket.org/ghentdatascience/ecmlpkdd20-papers/raw/master/ADS/sub_1006.pdf

Goes, F. R., Kempe, M., Meerhoff, L. A., & Lemmink, K. A. P. M. (2019). Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data*, 7(1), 57–70. https://doi.org/10.1089/big.2018.0067

Gómez, M. Á., Lago-Peñas, C., & Pollard, R. (2013). Situational Variables. In T. McGarry, P. O'Donoghue, & J. Sampaio (Eds.), *Routledge Handbook of Sports Performance Analysis* (pp. 277–287). https://doi.org/10.4324/9780203806913

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). Searching for a Unique Style in Soccer. 2014 *KDD Workshop on Large-Scale Sports Analytics*. arxiv:1409.0308

Hristovski, R., & Balagué, N. (2020). Theory of Cooperative-Competitive Intelligence: Principles, Research Directions, and Applications. *Frontiers in Psychology*, 11, 2220. https://doi.org/10.3389/fpsyg.2020.02220

Linke, D., Link, D., & Lames, M. (2018). Validation of Electronic Performance and Tracking Systems (EPTS) Under Field Conditions. *PLoS ONE*, 13(7). https://doi.org/10.1371/journal.pone.0199519

Pol, R., Balagué, N., Ric, A., Torrents, C., Kiely, J., & Hristovski, R. (2020). Training or Synergizing? Complex Systems Principles Change the Understanding of Sport Processes. *Sports Medicine - Open*, 6(1), 1–13. https://doi.org/10.1186/s40798-020-00256-9

Ric, A., Torrents, C., Gonçalves, B., Sampaio, J., & Hristovski, R. (2016). Soft-Assembled Multilevel Dynamics of Tactical Behaviors in Soccer. *Frontiers in Psychology*, 7. https://doi.org/10.3389/fpsyg.2016.01513

1

Always think before computing!

— Rafel Pol[1,2] & Natàlia Balagué[3]

1: Real Federación Española de Fútbol (RFEF)

2: Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), University of Lleida, Spain

3: Complex Systems in Sport Research Group, Institut Nacional d'Educació Física de Catalunya (INEFC), University of Barcelona, Spain

INTRODUCTION

What could we tell analysts to help them better connect theory and practice? Let's start from the beginning: why do we analyse the game?

R: There are three types of analysis that practically all professional clubs apply today. The first is oriented toward the scouting of talented players that the club may want to sign. This entails analysing the market to gain insight and decide whether or not players might be an attractive choice for certain positions. The second is analysis of opponents (teams or players), which is aimed at understanding the characteristics and weaknesses of rivals to prepare for competition against them. The third is player performance analysis, which is currently focused on kinematic variables (e.g. total distance covered, velocity, etc.). I am not aware of any elite club in the world that does not use this level of analysis.

N: This means that there are three organisational levels (club, team and player) that are involved in such performance analysis. Any decision made at a higher level (club), such as signing new players, affect those made at lower levels (e.g. team strategies, players actions) and vice versa. That is, the composition of the roster will impact playing systems and competitive strategies, which, in turn, bind the performance and actions of individual players (Balagué et al. 2019); Pol et al. 2020). Likewise, lower-level events (e.g. an injured or underperforming player) shape higher-level ones (change of playing strategies, signing of new players). It is important to take into account that these three levels cannot

be understood in isolation when thinking about performance analysis.

R: In fact, these three levels share a common limitation, inherited from an overly reductionist brand of sports science: they underestimate context.

a. Club level: When signing players, we often forget that some of them perform better in certain clubs and worse in others. Big data and statistics are used to confirm that a player is good at performing certain actions; it is then inferred that s/he will perform other actions equally well. But we must also examine the context in which s/he is performing these actions exceptionally well. There are also other areas outside the context of the match that allow us to predict whether or not a player will adapt well to the club. A fast and skilled dribbler may be invaluable in a team that covers a lot of distance due to its playing style. However, if we put that same player in a team that spends a lot of time in possession of the ball and that generally forces the opponent to retreat, s/he will lose that running space and, suddenly, no longer seem like such a skilled dribbler. Likewise, we might say that a player is the best centre-back without specifying that his or her team plays close to its own goal, leaving little space behind them. This player might be very good at confrontation but might not be a very fast runner. If s/he is signed by a team that plays in possession of the ball in the opponent's half, s/he will be forced to run long distances, face multiple players at once in a larger space and enjoy fewer one-on-one challenges in her/his territory. As a result, he or she won't just lose her/his reputation as the best centre-back but will come off as a shoddy defender. In contrast, consider a fast and small centre-back, such as Javier Mascherano, who played with Barça. In the context of that team, he was undoubtedly one of the best centre-backs in the world. However, he probably would have had more difficulty performing at that level had he been on a team that played defence closer to its own goal. In fact, when playing with Argentina, which does indeed play defence closer to their own goal, he was more of a midfielder than a centre-back. Players who are deemed the best in their position may end up seeming second-rate when the context is changed.

b. Team: When analysing an opponent, there is a general tendency to make statements such as "this team usually presses in this way" or "they play with the ball like this". The idea is to offer an average, or a general rule, of how the team plays. But "this is how they press" can only be confirmed when the opponent does not force them to play differently. For example, this statement will no longer hold true when playing against a team with the resources to resist such pressure. Identifying an "average" or a "summary" in this way does not provide information about what a team will do in a context in which their standard playing style is no longer effective. This analysis, therefore, becomes useless if, as a team, we manage to overcome our rival. And since that is the general objective, it is useless all around. Nevertheless, averages are compiled for teams to identify their characteristics and label them, despite the fact that these characteristics will disappear as soon as the context of the game changes.

c. Player: When analysing the performance of individual players, we once again seek out these averages and compare them with the team average, sometimes with the player averages and sometimes with what is typical in that level of competition. These averages offer little information about whether or not a player is capable of adapting to the demands arising from the context of the match.

At all three levels of analysis there is an attempt to extract quantitative data and little emphasis on the context.

Absolute, decontextualised quantification is very uninformative. For example, the simplest way to objectify perceived exertion (PE) is with the Borg scale (RPE 6-20 or CR-10) (Borg, 1998). Based on my experience in different situations and with different teams, the information obtained from PE ratings means very little if the players do not value this type of monitoring and if it is not adequately contextualised. We must encourage and harness athletes' ability to perceive and integrate information and put it at the service of monitoring. These are exceptional means of injury prevention and promote autonomy and self-management of workloads (Pol et al. 2018).

We can gather a great deal of information through conversations with athletes about how they feel in similar situations. This also may help the player become more aware of her/his condition and learn to express it as a comparison to her/himself and not to an external scale. But I think that if I wanted to publish these data I would be required to quantify them based on an external universal reference to objectify the differences between performance on different days and across different weeks. That would be difficult to publish, don't you think, Natalia?

N: Right, sports sciences place a greater emphasis on quantitative research, easier to objectify, and relies too heavily on group data averages.

However, it is possible to operationalise psychological or sociological information extracted from interviews and focus on individual data obtained from time series (Molenaar, 2004; Nesselroade and Molenaar, 2010: Rose et al. 2013). The main experimental model used to test research hypotheses in sports sciences is the comparison of group data means through inferential statistics. The fact that the averages obtained do not correspond to any specific individual is generally not taken into account when interpreting and applying research results. In fact, neglecting the idiosyncrasy of the structure of time-dependent variations within a single individual and assuming the equivalence between inter- and intra-individual variability is a common mistake of this type of research that leads to erroneous interpretations and inadequate practical applications (Rose, 2016, Balagué et al. in press).

Ignoring the context in which complex adaptive systems interact can be very misleading. In particular, when such context is non-stationary and fast changing, as during sports games. For example, the total distance covered by a player or her/his running velocity may be functional or un-functional depending on the context. Since every competition and every rival is different, it would be a mistake to compare kinematic information, obtained either at individual or collective level, without an adequate contextualisation and interpretation. In this sense, examining the dynamics of change of the variables under study is more informative to recognise whether performance stagnation or deterioration is taking place. This is applicable to all the aforementioned levels and requires updated assumptions and deep theoretical understanding of the examined processes.

R: Most clubs have entire departments devoted to generating analysis tools that, in practice, add no value to the work of the coaches. The problem lies in developing applications or reports without first thinking about the underlying theoretical assumptions. If the phenomena to be analysed are not properly understood, it is difficult to create effective applications. The first thing to do is to improve our understanding about the principles that lead the behaviour of the systems we deal with: the game, the team, the players. Only then a right answer to the question "what do we need to analyse?" will emerge. With the current approach, countless resources are being wasted.

N: Years ago, we published an article titled "Thinking Before Computing: Changing Perspectives in Sport Performance," (Balagué y Torrents, 2005). It might be useful to look back on its message to clarify what it means performance analysis from a complex perspective. Certain study variables such as entropy, big data or more advanced processing techniques (e.g. deep learning) are generally associated with the analysis of the game as a complex phenomenon. However, processing large volumes of data and variables compiled from these techniques is of little use if there is not accompanied by a deep understanding of the process under study. It must also be noted that merely processing common variables in the study of complex systems does not necessarily mean that a complex analysis of the game is performed. For example, the entropy measures, usually associated with the study of complex systems, can be used to answer a very simple question about cause and effect relationships that has nothing to do with a complex perspective. Essentially, in this perspective, we are not just changing the type of data analysis or study variables but the type of questions, assumptions and research interpretations. This is something we need to emphasise so that resources are not wasted. Otherwise, the same mistake will be made over and over, just with more sophisticated technology and more data or new variables, but still without improving the understanding of the phenomena

> **“Statistics itself, which does have clear limitations, is not the main problem”**

under study.

R: In fact, this lack of reasoning is more common in science than you might think. The journal Nature recently published an article signed by 850 scholars of different academic ranks that denounces the misuse of statistics in research. Specifically, it calls for the retirement of statistical significance to put an end to publications that report “hyped” results and dismiss crucial effects (Amrhein et al. 2019). Even the American Statistical Association has spoken out against the misuse of inferential statistics and its role in promoting erroneous and biased findings (Wasserstein et al. 2016; McShane et al. 2019). This tendency is not exclusive to performance analysis, but pervasive throughout sports science.

N: Statistics itself, which does have clear limitations, is not the main problem. It is the misuse and misinterpretation of statistics in an oversimplified research context. There is a general belief that exceeding a threshold for statistical significance means a result is real (Wasserstein et al. 2016; Amrhein et al. 2019, McShane et al. 2019). Without a deep understanding of the assumptions behind the research hypotheses, methodological decisions, data analysis and interpretation, the research is no longer practical but rather may add to the confusion and harm the advance of knowledge. If we do not understand the theory, we cannot propose useful practical applications. We need to be aware of the norm in sports science: isolating variables, comparing group averages, obtaining results based on statistical significance and putting them into practice as general truths. All this represents a significant bias for the advancement of knowledge. The reductionist tendency, inherited from medicine and psychology, and over-simplified theoretical assumptions limit the evolution of our understanding of complex phenomena linked to sports performance.

It is possible, however, to turn away from tradition and opt for other models, such as those of complex systems, widely applied in several emerging branches of biology. Of course, this option does mean overcoming the resistance and difficulties that come with any paradigm shift. We now find ourselves in the midst of that transition, forcing models that are to some extent incompatible to coexist. However, the difficulties associated with such paradigm shifts should not discourage the pioneers because they are on the right track. Interdisciplinary work is being encouraged in many scientific branches and the contributions of physicists, mathematicians and computer scientists, collaborating with sports scientists, are key in performance analysis.

R: However, it is important to highlight that merely creating multidisciplinary teams will not overcome the obstacles of studying a phenomenon that belongs to no specific discipline: neither the game nor the player fit within these disciplines. Therefore, it is not just about adding new professionals and perspectives from different branches, but about sharing a common approach to how the systems that we work with function – whether players or teams – and making contributions from that common vision.

N: Right, there are transdisciplinary initiatives that pursue real knowledge integration, but multidisciplinarity is in fact non integrative (Balagué et al. 2017). It is necessary to clarify what a real integration means, and which type of approaches may satisfy it. Some authors suggest that training-methodology departments should integrate knowledge on the basis of an ecological dynamics approach (Rothwell et al. 2020). This perspective, however, focuses on the level of perception and action, while neglecting other relevant levels related to strength and conditioning (e.g., biochemical and physiological). In line with trans- or cross-disciplinarity, some authors sustain that integration requires a common scientific vocabulary shared across disciplines (Balagué et al. 2017); Hristovski et al. 2017). In this direction, Glazier (2017) proposes the unification of knowledge based on the concept of constraints and Balagué et al. (2017) suggest the Dynamic Systems Theory as an ideal integrative framework. DST does not limit the common vocabulary to the concept of constraints but to all the DST concepts; valid for all levels of analysis (see Fig. 1).

However, despite having powerful tools such as Dynamic Systems Theory, it is quite challenging to evaluate the dynamics of change in collaborative and competitive processes in which rivals never behave in exactly the same way and there are no static situations, as in football.

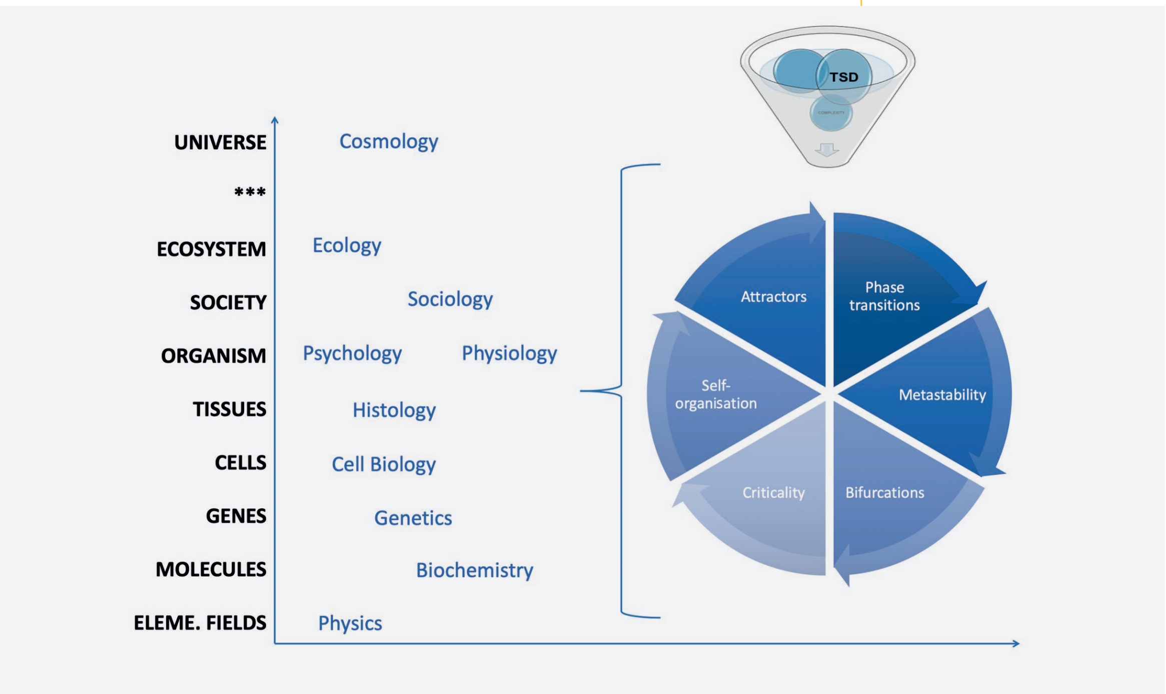
R: Perhaps we should change the type of question or the objective of performance analysis. If we have a deep understanding of the dynamic mechanisms of players/teams to adapt to the context of a match or training session, and develop them during training sessions, we could even forget about analysing the opponent. The issue is that training sessions are often focused on the rival, on their weaknesses and strengths as detected from analysing their game. In response, optimal, predetermined solutions are prepared. But if we changed the objective from having the team or player learn specific solutions to improving their potential for adaptation and diversity (Pol et al. 2020) and their ability to resolve unpredictable situations – such as those that happen regardless of how we train – we would not even need to analyse the opponent or, at least, not in the way we currently do. We could focus instead on analysing the state of our team and deciding whether we should generate new training challenges or a sense of confidence by strengthening scenarios already under control. But, of course, for that we would need the deep understanding of the theoretical foundations of performance that we mentioned earlier.

N: Even separating training and competition would no longer make sense. There is training in competition and competition in training, so why shouldn't we view them as continuous, adaptive processes with disturbances and constraints that are more or less intense and/or global?

R: In fact, training while competing is the best type of training; this becomes very clear in teams that play every three days. One of the main problems that usually crop up after a team that is not accustomed to competing at a high frequency qualifies for a European competition, is that the coaches underestimate that they can no longer train as much as they used to. If they do not address the process as a whole, understand that competition is part of the process, and eliminate a significant part of the training, they end up overwhelming the team. Even if the coach is changed, it becomes very difficult to revive the team. This situation repeats itself every year.

N: There is a fundamental problem in our understanding about the systems we deal with: it is generally assumed that they are dominated by

Figure 1.
Structure and organisation of scientific fields sharing the same concepts and principles from Dynamic Systems Theory.
v



> **More information does not mean more knowledge**

the dynamics of their components, instead of being dominated by the dynamics of the interactions among such components, as occurs in complex systems (Delignières and Marmelat, 2012). This explains why training focuses on components and underestimates the counterproductive effects this may have on competition. For example, fragmenting and stabilising excessively positional play in attack and defence (separately) during trainings may harm the transitions during competitions. Moreover, there is a crucial training goal in team sports that is often underestimated by this component-dominant approach: developing interpersonal synergies among players. These synergies, which are the basis of the team's functional diversity potential and competitiveness (Pol et al. 2020), require collective challenges and constant interaction between players to overcome them. If training objectives change their focus, the line between training and competition will be blurred. Technology must serve that understanding, not the other way around. Different departments of the club, even members of the coaching staff, often analyse competition and training separately. That is, the performance department and the physical trainers focus on analysing players and their physical and physiological response during training through GPS data, and analysis departments focus on the competition (event data and movement tracking) through video analysis. Perhaps this fragmentation should also be put into question. These common and artificially built barriers can be reduced under the proposed paradigm shift, and specifically, under the umbrella offered by the Dynamic Systems Theory.

R: The key to knowledge integration is a shared understanding of the phenomenon to be studied. You can have a team comprising professionals from different disciplines working on an integrated job, but if the biomechanist tells you that a player missed a pass because s/he shot from the wrong angle, while the physical trainer says that s/he was worn out by the end of the game because s/he played for too many minutes and did not rest enough, and the physiologist agrees with the physical trainer and adds that the results of the last blood analysis were not good enough, what do you do? You say you are doing multidisciplinary and integrated work because every specialist provides their view on the phenomenon, but if all these specialists understand it differently, and look at it from their own perspective, it becomes very difficult to generate new knowledge. In such cases, a multidisciplinary team is of little or no use and may even detract from the work. If we do not change our understanding, any other efforts we make are useless.

N: More information does not mean more knowledge. Scientific specialisation has provided a lot of detailed information, but this does not necessarily imply greater comprehension and knowledge. For example, specialisation does not encompass the fact that the levels of analysis previously mentioned – club, team and player – do not exist independently of each other but are related by circular causality (Balagué et al. 2019). In cooperative and competitive environments, maintaining and developing the potential for functional diversity/unpredictability – also defined as intelligent cooperative-competitive behaviour – at different levels is crucial (Hristovski and Balagué, 2020). In other words, having a smart club, a smart team, and smart players means evading and escaping situations of reduced possibilities of goal achievement in different contexts. Opponents try to reduce that diversity potential intervening at all levels, and the goal is to maintain and develop it despite these disruptions. Functional players are those that contribute to the collective goal, and functional teams are those that escape and quickly regain the diversity potential despite their opponents' actions. One fundamental advantage is that this potential can be measured using a common variable: information entropy.

R: In conclusion, to improve the methodologies of game analysis we must start by improving our understanding of the phenomena under study. When our understanding changes, the questions we ask ourselves change; and when the questions change, the way we analyse phenomena necessarily changes as well.

REFERENCES

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists Rise Up Against Statistical Significance. *Nature*, 567, 305-307. https://doi.org/10.1038/d41586-019-00857-9

Balagué, N. y Torrents, C. (2005). Thinking Before Computing: Changing Perspectives in Sport Performance. *International Journal of Computer Science in Sport*, 4, 2, 5-13.

Balagué, N., Torrents, C., Hristovski, R., Kelso, JAS. (2017). Sport Science Integration: An Evolutionary Synthesis. *European Journal of Sport Science*, 17, 1, 51-62. https://doi.org/10.1080/17461391.2016.1198422

Balagué, N., Pol, R., Hristovski, R., Torrents, C., Ric, A. (2019). On The Relatedness and Nestedness of Constraints. *Sports Medicine Open*, 5, 6. https://doi.org/10.1186/s40798-019-0178-z

Balagué, N., Hristovski, R., Almarcha, M. C., García-Retortillo, S., Ivanov, P. (in press) Network Physiology of Exercise: Vision and Perspectives. *Frontiers in Physiology*.

Borg, G. A. (1998). Borg's Perceived Exertion and Pain Scales. Champaign (IL): *Human Kinetics*.

Delignieres, D., and Marmelat, V. (2012). Fractal Fluctuations and Complexity: Current Debates and Future Challenges. *Crit. Rev. Biomed.* Eng. 40, 485–500. https://doi.org/10.1615/CritRevBiomedEng.2013006727

Glazier P. (2017). Towards a Grand Unifying Theory of Sports Performance. *Hum. Mov. Sci.*, 56:139–56. https://doi.org/10.1016/j.humov.2015.08.001

Hristovski, R., Aceski, A., Balagué, N., Seifert, L., Tufekcievski, A., Aguirre, C. (2017). Structure and Dynamics of European Sports Science Textual Contents: Analysis Of ECSS Abstracts (1996–2014). *European Journal of Sport Science*, 17, 1, 19-29. http://dx.doi.org/10.1080/17461391.2016.1207709

Hristovski, R., Balagué, N. (2020) Theory of Cooperative-Competitive Intelligence: Principles, Research Directions and Applications. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.02220

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *American Statistician*, 73(sup1), 235–245. https://doi.org/10.1080/00031305.2018.1527253).

Molenaar, P. C. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement*, 2(4): 201-218.

Nesselroade, J. R., Molenaar, P. C. (2010). Analyzing Intra-Person Variation: Hybridizing the ACE Model with P-Technique Factor Analysis and the Idiographic Filter. *Behavior Genetics*, 40(6), 776-783.

Rose, T. (2016). *The End of Average: How to Succeed In a World That Values Sameness*. USA: HarperCollins

Rothwell, M., Davids, K., Stone, J. A., O'Sullivan, M., Vaughan, J., Newcombe, D.J., Shuttleworth, R. (2020). A Department of Methodology Can Coordinate Transdisciplinary Sport Science Support. *Journal of Expertise*, 3, 1, 55-65. http://shura.shu.ac.uk/25980/

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *American Statistician*, 70(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

2

Unifying player IDs for the future of football analytics

— Hans Thies[1], Bahaeddine El Fakir[1] & Nicolas Evans[1]

1: FIFA

INTRODUCTION

What may have started out as a trivial problem some years ago when the technology landscape was only starting within sports and football in particular has become one of the main challenges for all people involved in the data value chain in the sport: collecting data and allocating it to the correct player while simultaneously providing access to those who are entitled to it.

While single providers had, and for larger clients continue to have, solutions whereby each player can be set up individually, parametrised and displayed in a customised way, for the vast majority of other users this luxury does not exist and has become a problem. So much so that it can be make or break for the use of technology.

The core stakeholders are performance coaching staff (as proxy for the players) who need to efficiently process many data sources to inform training and game decisions; recruitment and scouting staff who increasingly rely on data as a factor in talent identification; and the broader public, via media channels that look to tell a more data-driven story of the game. Their task is becoming harder because of the multitude of service providers for comparable data collection (such as positional tracking and event data), as well as increasing new individual wearables and innovations used at home or in training alongside a multiplication of tournaments staged by different competition organisers. This has created an almost unnavigable web of data allocation to an individual subject, meaning that each stakeholder needs to use different platforms and that data can only be harmonised through massive efforts by sports science teams.

The size of the task at hand for creating a global identifier for football players has only become apparent thanks to work that was started as part of a much larger project involving the more fundamental task of player registration. As FIFA increasingly dealt with the potential of technology for developing the game and the football experience – and began conceptualising its football data ecosystem – one previously developed and readily available solution was found to lend itself almost perfectly to the aforementioned problem of being a universal identifier for individual players: the FIFA Connect ID.

This chapter will outline in detail the origin of the FIFA Connect ID, with the numerous legal and technical challenges it was designed to solve. It will also elaborate how this key piece of the football data ecosystem will be able to simplify data acquisition and sharing alongside data format standards, while being compliant with demanding data protection regulations.

THE STARTING POINT OF A GLOBAL UNIQUE IDENTIFICATION SYSTEM: ELECTRONIC STAKEHOLDER REGISTRATION IN FOOTBALL

FIFA, as the worldwide governing body of football, has made it its core mission to “make football truly global”, as elaborated by President Gianni Infantino in the strategy document Vision 2020-2023 (FIFA 2020b). The FIFA Connect Programme was born with the goal of achieving just that in the domain of stakeholder registration – developing the core administrative task that enables member associations to run and organise their domestic competitions, which in turn are at the heart of every member association’s mission. It can therefore be considered a digital football development programme aiming at comprehensively solving the issues around registrations of players and officials at a global level. In 2014, the goals of the programme were defined. Based on an analysis of the system landscape, four essential challenges were identified in the area of stakeholder registration:

First, more than half of FIFA’s 211 member associations did not use an electronic registration system but instead relied on manual processes, with the obvious associated downsides in efficiency and data management. While this offered a carte blanche for new development, it highlighted the size of the task at hand for rolling out global electronic player registration. Second, around an additional 20% of FIFA’s member associations were using outdated legacy systems, which did not enable them to reap the full benefits of registration data (more details on this below). In addition, many of these electronic registration systems were often limited to top-level competitions thus entailing the dual task of adapting and expanding at the same time. The third identified difficulty was that, in many cases, irrespective of how data was collected, it was often located in multiple, unconnected data silos within the member association. Linking these silos while integrating a new system would be additional work. Last, the data itself that was being collected & processed within stakeholder registration was extremely heterogeneous. As a result, it was not easily possible for member associations to exchange data with other institutions, integrate data from external sources, or even change their existing supplier(s) due to rigid formats and unique management systems.

Bearing these different sets of prerequisites in mind, it must be noted that it was in no way FIFA’s mandate or desire to dictate the exact details of stakeholder registration, including which systems to use, or to force member associations to switch systems. It is worth mentioning that some member associations

had already developed truly state-of-the-art systems that could not be significantly improved upon for domestic registration purposes, even by taking a centralised development approach. In order to account for this, FIFA made it its mission to help its members to overcome the aforementioned challenges, using a tailor-made approach that did justice to their different development stages. From member associations using paper or spreadsheet-based solutions, through those that use legacy-based systems in need of modernisation, to some using best-in-breed solutions, FIFA wanted and continues to support development and implementation in a manner that is cost-efficient and takes the best possible care of the funds intended to develop football and making it truly global. Consequently, the following goals defining the vision of the FIFA Connect Programme were outlined:

- All 211 member associations using an electronic system for stakeholder registration and competition management: the use of this system should gradually be extended down within the pyramid of football registration (from international and national competitions down to more scattered regional and local organisations), to reflect all clubs, players, referees and coaches taking part in organised football within the country. This meant introducing for the first time such a system in around 50% of FIFA's member associations.
- Standardisation of processes and data: all member associations should adhere to best-practice standards with regards to the data fields (and format) collected, the data entry, duplicate handling, as well as the data validation/approval, as a pre-condition to raise the quality of data, and making data interchangeable. While it sounds straightforward, the challenge would be to agree on this format and ensure utmost compatibility with the different existing systems.
- Secure identification of football stakeholders, on a global level: one globally valid unique ID should enable matching of data with regards to stakeholders – from different systems, member associations and contexts – to enable exchange and enriching of data and a 360 degree view on an individual's career in football. A standardised communication infrastructure should enable the flow of data between different organisations and systems.

In order to best assist member associations to adjust to the exact situation and capabilities, a modular approach was consequentially developed and embedded in the FIFA FORWARD Programme (FIFA's development programme for its member associations), which provides the necessary capital to finance infrastructure, hardware and complementary projects to raise the standard of every member association accordingly. The five pillars of the programme that are depicted in Figure 1 are:

1) The FIFA Connect Processes: a library of best-practice registration processes for stakeholders, created by an expert group including member association staff and business experts. The processes are offered as a service to member associations and registration system providers, to ensure adherence with the best practices in the domain.

2) The FIFA Connect Data Standard (FIFA 2020a): this specifies the data model to define the basic properties of entities and to exchange football-related data within the FIFA family and other interested parties. The purpose of this specification is to support them in building applications and processes that enable the exchange of core football data – such as clubs, player registrations or match reports – in a defined and standardised way. It was defined, and is subsequently revised and improved by, a steering board comprising subject-matter experts from member associations, confederations and software suppliers led by FIFA.

3) The FIFA Connect Platform: a standalone software module allowing member associations to jump-start electronic stakeholder registration. The system is configured according to the FIFA Connect Data Standard & Processes, based on the specific requirements and setup of each individual member association using it.

4) The FIFA Connect Competition Management System: a software module building on top of the FIFA Connect Platform, enabling member associations to organise their competitions end-to-end fully digitally, and totally paperless; thereby enhancing registration and competition data to a more complete dataset of each stakeholder and his or her activities within football.

5) The FIFA Connect ID Service: a web- service that can be connected to any registration system in the world, assigning a globally valid unique identifier (the FIFA Connect ID) to stakeholders registered in a decentralised way within the individual member associations system, taking care of double-registrations (duplications), and providing a communication infrastructure for connected systems to exchange football-related data associated with entities that have been assigned a FIFA Connect ID, based on the FIFA Connect Data Standard.

With the development objective of making this suite of tools and funding available to the member associations, the implementation then comes down to the tailor-made project plans to allow a future-proof stakeholder registration system to be put in place. This will achieve the first pillar of having a digital system in place that is capable of performing the required task.

Figure 1.
The FIFA Connect Programme as a digital football development programme "off the pitch"
v

IMPLEMENTING THE FIFA CONNECT ID SERVICE: HOW TO GO ABOUT SECURELY AND COMPLIANTLY IDENTIFYING FOOTBALL STAKEHOLDERS ON A GLOBAL LEVEL?

Having outlined the historical raison d'être of the FIFA Connect ID, the remainder of this chapter will focus on the FIFA Connect ID Service. It will look at the challenges during its development as well as its general principles and benefits, with a special focus on how it may now serve the performance data and analytics community in ways that could not have been imagined at conception. In particular, it will explore the experience of overcoming the paradox of needing to create a system that is as thorough as possible (to avoid duplicates and errors) but which at the same time is as minimalistic as possible in order to respect myriad data protection standards.

The problem of uniquely identifying a person is of course not a new one. Most countries assign a unique ID to their citizens, either upon birth, by issuing an ID document or passport, or through entry into the social security system. Similarly, companies assign unique identifiers to their customers and partners, mostly in a CRM (Customer Relationship Management) system, to map data from different sources and systems and get a better view of customer behaviour and the most appropriate strategies to improve revenues and retention.

The goal of the FIFA Connect ID project is to provide every football stakeholder taking part in organised football anywhere in the world with a unique identifier: the FIFA Connect ID. This identifier, as explained in the remainder of this article, is supposed to serve the purpose of identifying the player (and football officials, as well as clubs and stadiums) in football management systems in the context of a wide range of scenarios in football, including (but not limited to):

- Registrations & transfers of players and officials
- Solidarity and training compensations for clubs
- Disciplinary matters
- Competitions and tournaments at domestic and international level (e.g. registration, accreditation, line-ups and match reports)
- Exchanging data between entities and systems (youth football to professional football or between a national league and member association for example)
- Enriching the data (e.g. with tracking & performance data), combining data from different data sources

The important part is to provide the identification of the stakeholders in a globally valid, standardised and – most importantly – machine-readable way. During the design of the system, the two most prevalent constraints became apparent:

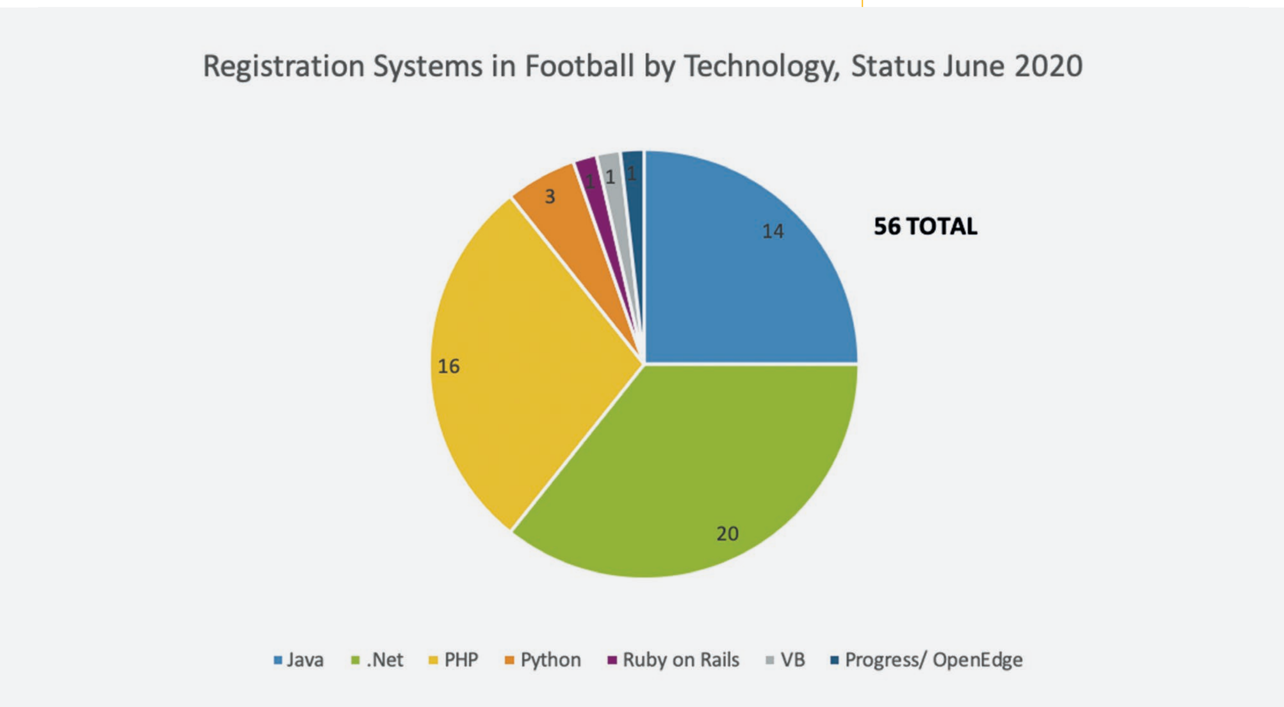
1) HETEROGENEOUS REGISTRATION SYSTEM LANDSCAPE:

As mentioned earlier, many of the FIFA member associations were still using paper- or spreadsheet-based solutions for stakeholder registration. FIFA has addressed this challenge by offering its member associations a free cloud-based registration system, fully configurable to the member association's needs. In addition, many member associations were using outdated legacy systems based on a wide variety of technologies, processes, and data. FIFA decided to address this by offering the FIFA Connect Data Standard, and the FIFA Connect Processes as tools to unify, standardise, & establish best practices in electronic stakeholder registration worldwide. However even after a consolidation of registration systems initiated by FIFA to introduce an electronic registration system in all 211 member associations, replace legacy systems with the FIFA Connect Platform, or renew them with more appropriate systems following the minimum standards, this still left the challenge of integrating a heterogeneous system landscape of systems using different technologies with a central service assigning the FIFA Connect ID. A current assessment of the varying systems in use for stakeholder registration in the 211 member associations, and the corresponding technology used, is shown in Figure 2.

2) NUMEROUS USE CASES VS. A WIDE VARIETY OF LOCAL LEGISLATION, IN PARTICULAR WITH REGARDS TO DATA PROTECTION:

While on one hand, the clear goal of the system is to provide a globally valid, unique identifier to stakeholders, it is on the other hand also FIFA's goal to collect as little data as possible to comply with the principles of data privacy by design (Cavoukian, 2010) and make it clear to everyone involved that FIFA advocates that the data and the associated (financial) benefits that can be derived from it belongs to the players and the member associations. These two goals or principles are diametrically opposed to each other: the more (specifically sensible) data is collected centrally, the more certain the system can be that the identifier is unique and duplicates are avoided. If the system were to collect e.g. biometric data, (email) addresses, clear text names and date of birth, national identifier numbers or passport numbers, etc., the unique identification could be provided almost automatically and close to perfectly, but the principles of data minimisation and proportionality would be violated. In addition to being contrary to FIFA's goal of storing as little data centrally as

Figure 2.
Assessment of registration systems in football by technology, status June 2020
v



possible, it certainly would not have been in accordance with the General Data Protection Regulation (GDPR) legislation in Europe as introduced in May 25th 2018 (European Union, 2020), just to give an example, even if it may have conformed to other local laws. However, since the FIFA Connect ID service needs to comply not only with European legislation, but the data protection legislation in the territories of 211 member associations, it must in essence meet every existing data protection standards in the world. The means to address this challenge was to identify the strictest data protection legislations and principles and base the overall system architecture on these while addressing the "data privacy by design" framework.

The establishment of the FIFA Connect ID Service therefore hinges on finely balancing the amount of data required to identify the player and other stakeholders while not only safeguarding personal data in the process but also ensuring that the global system can function for all its required objectives, ranging from player registration to competition management, international transfers or club compensation from youth player development. These are the conceptual specifications that the system architects were facing.

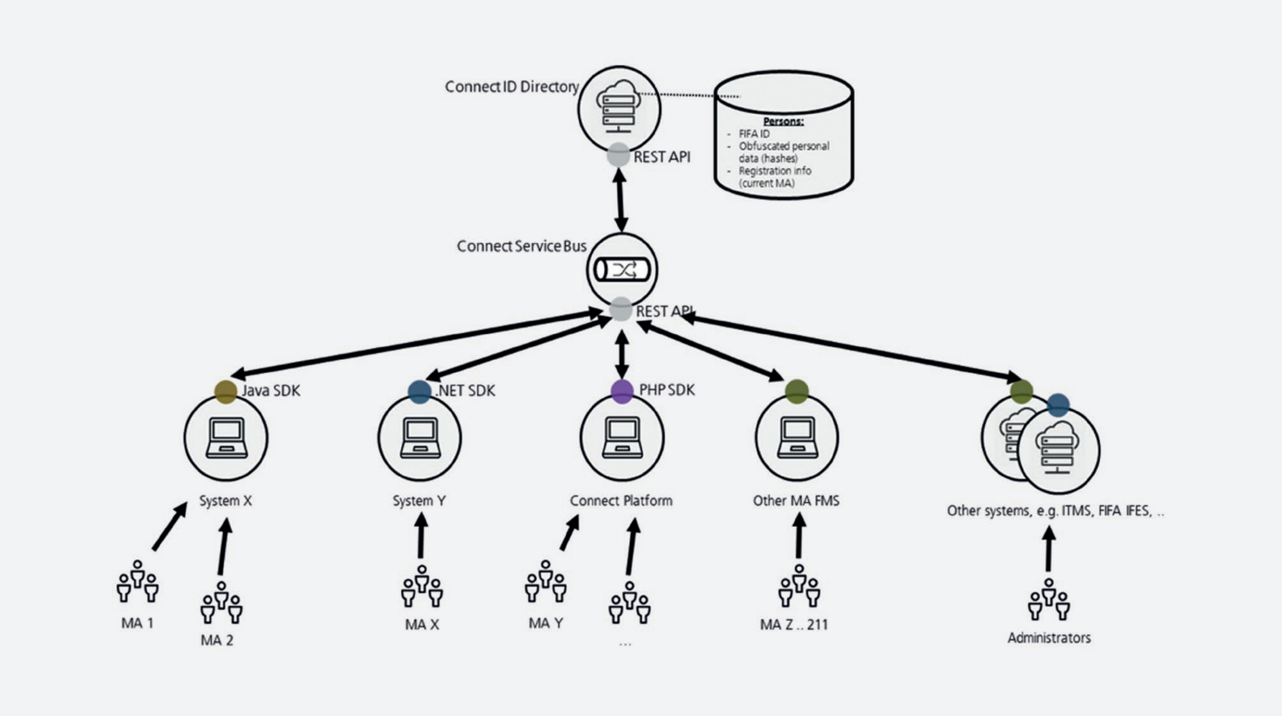
DESIGNING A FUTURE-PROOF AND COMPLIANT ARCHITECTURE FOR THE FIFA CONNECT ID

With the restrictions set by the legal framework, the ambitious task of a global identifier system needed to be designed to allow for maximum functionality while respecting individual privacy. As a consequence, the architecture of the Connect ID system needed to fulfil three main objectives:

First, the mapping of stakeholders through a unique identifier in a multitude of heterogeneous systems. These systems include registrations systems of FIFA member associations, competition management systems of confederations, FIFA's own systems (such Transfer Matching System which governs international player transfers or FIFA's Internal Football and Event management System (IFES). In addition, the mapping needs to take further systems into account as the scope is being extended, for example to include the CMS of media companies and data providers. Second, as part of the assignment of the unique identifier (the FIFA Connect ID), the systems needs to be able to reliably identify duplicates. This includes, if necessary, identifying different languages, scripts (such as Latin, Arabic, Cyrillic, Chinese, Japanese, Thai, etc.), considering different name variants and name sequence conventions, as well as identifying errors during data entry, both due to fraudulent behaviour and mistakes. Third, the system needs to provide an infrastructure for the connected systems to communicate with each other in a standardised way. If all systems in the ecosystem needed to communicate with each other, the amount of required interfaces in case of simple 1-1 APIs would be almost infinite following Gauss' (n)(n-1)/2, with >200 systems in scope. This is why a standardised interface, with a uniform data standard is so essential to achieve the overall objectives of the Connect ID Service.

From a non-functional perspective, the most important requirements are those of performance (e.g. speed and throughput of the system), availability, and scalability, with more than 200 systems and in excess of 100 million potential stakeholders in scope. This furthermore needed to be achieved with the two previously mentioned principles of *privacy by design* – the general design whereby technical choices are made assuming that the data belongs to the data subjects and that their data should only be processed to the extent absolutely necessary – as well as simplicity of access to the infrastructure (and the Connect ID Service to assign the unique identifiers itself) allowing for the widest possible range of users to take advantage without discrimination.

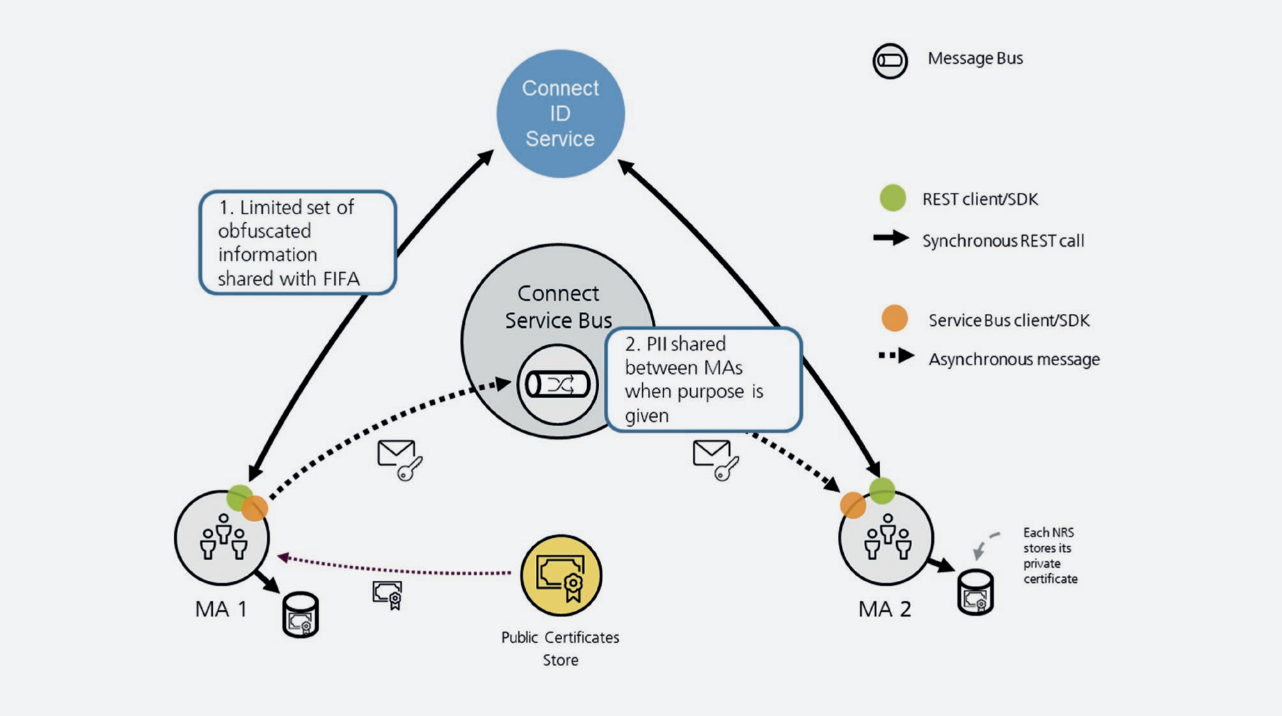
The key means to achieve the first principle followed during the architecture design were decentralisation, data minimisation, and data pseudonymisation. Instead of designing the system and afterwards trying to accommodate data protection, it was designed with those principles in mind. A working group including FIFA and research partners from industry and universities defined the system architecture following a decentralised architecture: while the data about the stakeholders was to remain in each member association's Football Management System, where it would be made accessible for when a concrete purpose for the processing of the data is given, the Connect ID Service was only to store the minimum set of data centrally. The minimum set of data to store was identified in a specific project phase, the standardisation prototype, experimenting with different combinations/sets of data. Most importantly, the prototype showed that a pseudonymisation of the stakeholder data could be achieved by not storing stakeholder data in clear text, but in hashed form, by pre-calculating several hashes representing several combinations of the stakeholder's data in a standardised way. By using underlying name databases in different languages, transcribing and translating the data, and then applying machine learning, the name can be transformed into several possible standardised variants, e.g. the inputs "Dick", "Richad", "Ричард", "Richard" can all be standardised into a variant "Richard" and afterwards hashed using a secure algorithm



^
Figure 3.
High-level architecture of the FIFA Connect ID Service

(SHA-512), thus allowing both for a pseudonymisation of the data AND a reliable detection of possible duplicates. The inputs of the systems can furthermore be used to train the underlying algorithm. Data retention and deletion policies were also embedded into the system and the appropriate mechanisms provided to the member associations to update the data accordingly.

The key means to allow for easy access to the infrastructure and the Connect ID Service were the use of standard web technology, such as REST APIs, to consume all services (both external and internal support services) and the use of SDKs, which allow the users to integrate all services through native methods. Based on the analysis done as shown above, in excess of 90% of the football management systems and 96% of the member associations were using systems in Java, PHP, or .NET. As a consequence, the SDKs were developed in these languages. In case of the other languages in use, an integration through the REST APIs or through a proxy application using the SDKs can be developed. The SDKs make integration significantly easier, most importantly by encapsulating each different process to be integrated with the Connect ID Service in one simple native method. In addition, they provide the additional benefit by validating incoming and outgoing messages, and providing methods for creating classes in the FIFA Connect Data Standard (e.g. POJO classes for JAVA). In this way, possible misunderstandings in the data transmitted to the FIFA Connect ID Service but, more importantly, between the different systems and member associations, can be ruled out in an elegant and simple-to-integrate way. If each member association was to create and/or validate their own messages e.g. in XML or JSON format, edge cases where certain information is missing or added would be almost impossible to test or rule out, leading to errors that are extremely difficult to catch by conventional ways of testing, such as unit or integration testing. By including the creation and validation, but also the sending and receiving of the messages into the SDKs, a much more reliable and dependable way of exchange standard messages can be guaranteed. In architecture, as demanded by data protection and data minimisation. The high level architecture of the service is depicted in figure 3.



^
Figure 4.
FIFA Connect Service Bus

To furthermore provide the best possible protection of the data, it is being encrypted and the obfuscated data stored in highly secure infrastructure. For the exchange of data between systems, an end-end encrypted, asynchronous service bus communication is being used, with the following advantages:

- The data is not accessible by the system operator (e.g. FIFA) and therefore only the sender and receiver can see the data.
- Via general message properties and topics in the Service Bus, FIFA can still regulate the use of the infrastructure.
- The SDKs ensure the proper use and generation/ validation of the standard messages. The member association can however also use the existing infrastructure to send custom messages, when the effort to create message generator/ handler in the SDK is not justified (e.g. for custom messages send only between two parties).
- The data is stored in the service bus until picked up by the recipient, thereby providing functionality and error-handling even in the case of a temporary outage of one of the systems.

Figure 4 illustrates the high-level process of the de-central assignment of a FIFA Connect ID. In a first step, the stakeholder is being registered in a member association, making use of the member association's own registration system, but following the FIFA Connect Data Standard and one of the 40 variations of the best-practice FIFA Connect Processes. The stakeholder data is obfuscated (hashed) and a limited set of obfuscated information is sent to the Connect ID Service in order to generate a FIFA Connect ID, which is sent back to the member association and stored locally in their own Football Management System. Only in case a clear purpose is given (such as a transfer of the player), PII (European Union, 2020) is being exchanged between the particular member associations in question. The communication between two systems through the service bus is always end-to-end encrypted using a pair of private (stored locally)/ public certificates (available in public certificate store) generated using a private key by each system provider, ensuring the data shared between two system operators is kept private and cannot be accessed by FIFA.

This comprehensive description of the system architecture shows the immense amount of thought that has gone into the process and has led to nothing less than a tool that is allowing 211 member associations from all sorts of sporting, economic and geographical contexts connect to a truly global system for the purpose of simplifying transactions for the whole game of football. With this groundwork done, it becomes apparent that a number of other use cases could hugely benefit from this unique ID system as will be explained below.

EXPANDING THE USE OF THIS UNIQUE ID FOR ALL DATA COLLECTION IN FOOTBALL

The FIFA Connect ID has been introduced as a mandatory aspect of stakeholder registration (§5.1 of the Regulations on the Status and Transfer of Players, FIFA (2020c) by now. As a consequence, more than 11 million FIFA Connect IDs in 185 member associations have been assigned to date (status July 2020), with an expected 30 million by end of the 2020. This wide of use of the FIFA Connect ID and its underlying infrastructure by FIFA, its member associations, and their system partners and suppliers, mean that the FIFA Connect ID and the FIFA Connect Data standard have already become a de-facto standard in the domain of registration and competition management in football. This is what makes the FIFA Connect ID an obvious choice going forward for all other data-gathering exercises such as EPTS, performance management and football analytics due to the ability to simply allocate data to the correct player. It equally means that players will be able to centrally collect any data associated with their FIFA Connect ID irrespective of which provider has collected it, what the competition is and if the player has been transferred.

The FIFA Connect ID Service can accordingly be opened to a supplier of performance-tracking data, software, or equipment. In particular, the following services can be offered to a provider of a third-party system:

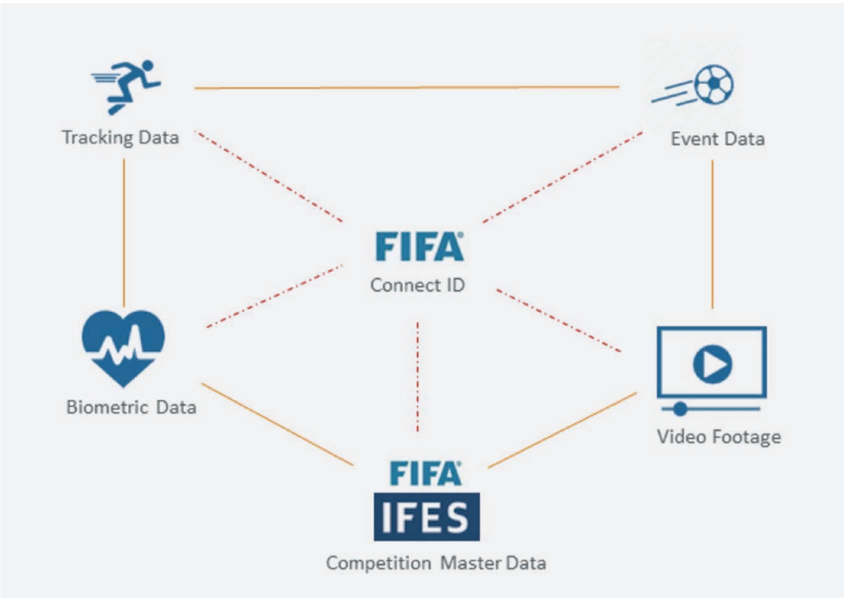
- Automatic assignment of FIFA Connect ID: Via one of the SDKs or the REST API, the provider's system may check the FIFA Connect ID of a player by providing the full name and date of birth of the player. This functionality will be restricted to professional players only and requires the legal agreement of the member association (and potentially other agreements) where the player is registered.

- Manual assignment of FIFA Connect ID: An alternative is for the tournament organiser to share the FIFA Connect ID of the participating players with its respective providers to ensure the officially collected data is correctly allocated. This is an approach which has already been successfully applied and tested in FIFA tournaments.
- Use of the FIFA Connect Service Bus: The FIFA Connect infrastructure can be opened for third parties to exchange generic and/or custom messages with member associations and FIFA systems. Similarly to the above case, the exchange of information will require legal agreements between the participating parties. Generation/validation of messages based.
- on the FIFA EPTS Standard Data format is currently not included in the SDKs, but can easily be included for the purpose of ensuring consistent message format & error handling.

With a number of technical options available already today, FIFA has started to collect data at its tournaments using the FIFA Connect ID as a means to allocate data from all of its performance data providers to the correct players. The synergies that were not necessarily foreseen at the beginning of the "electronic stakeholder registration" process are proving a most welcome option with the multiple data sources, formats and competitions collecting data, and are set to relieve many providers and team members from the arduous task of collecting, transforming and stacking the data from an increasing number of diverse sources.

As data is increasingly becoming the centre of attention in today's football landscape, more industry leaders are innovating and proposing new data sets to leverage. This will inevitably add to the complexity, the time consumption of ID mappings and numerous false positives. It is therefore of utmost importance to fine-tune the requirements to today's most demanding football stakeholders and start a shift towards using these unique identifiers to avoid kicking the can down the road and facing unsurmountable mountains of data that will become increasingly difficult to standardise, compare or even allocate correctly. Whereas the current solution of providing a unique reliable ID to an organisation's different data providers will only shift this cumbersome task from the contractor (the league or club) to the contracted (data providers), the FIFA Connect ID is planned to be accessible enough to data providers so that all parties – from member associations to data collectors – will be able to use it across all activities and therefore encouraged to use it as a global standard. In this sense the Football Data Ecosystem should provide a truly worldwide arena in which standardised data can exchanged without the need of moving between individual silos (from league to national team or following a player transfer for example).

A minor task that will be required to completely fulfil this need is for the FIFA connect ID to allow for IDs to be created for games and competitions and any potential other "stakeholder" required in football, which is one of the clear next steps.



<
Figure 5.
FIFA Football Data Ecosystem inputs using Connect ID

FUTURE CHALLENGES, CONSIDERATIONS AND NEXT STEPS

As stated above, the FIFA Connect ID became mandatory for any football stakeholder registrations in July 2020. While there are still some transitional problems to solve, on the face of it there is a solution that has not only gone through very careful due diligence but has also been designed with utmost consideration for a player-centric data system with ownership and control of the data resting with the actual stakeholder. As FIFA aims to make further headway in this respect, there are a few continued challenges that will be addressed in the coming months in order to fully onboard the technology providers:

First, the true rate of FIFA Connect ID roll out: while the first signs are hugely promising, roll-out will start in more advanced competitions and member associations meaning it may take a while for some younger players to be registered in the system. While this is most likely a matter of time, it is also worth considering the time it may take to enter all existing players into the FIFA Connect ID system in order to have a complete picture of the stakeholder landscape.

Second, making SDKs/REST APIs accessible publicly: while respecting data privacy, one challenge will be to make a player's FIFA Connect ID available to data providers who are collecting the data on behalf of the subject (or the subject's team). While there are plans to address this in the future, the exact mechanics will need to be worked out.

Lastly, interoperability of data formats and standards: while FIFA has addressed this topic for tracking data and is currently working on a global definitions manual for football events, it will take time for providers to adjust their data sets to be more uniformly usable. So, while the FIFA Connect ID will quickly be able to allocate the data to the correct user, it may still be a while before this user will be able to truly benefit from this data.

This chapter was about highlighting the immense amount of work that has gone into creating a truly global (and worthy of that name) unique identifying system for football stakeholders. While the primary purpose was more basic in nature – player and stakeholder registration – to allow for uniform match data collection, international transfers and player registration for tournaments, the FIFA Connect ID has proven an invaluable asset in the new football data ecosystem where allocation of collected data has been anything but a trivial task. Far from perfect and accomplished, the mission that FIFA set out to solve is however unique in its nature and has taken an approach that not only respects data protection on a global scale but is, arguably, future-proof. It will help solve many issues that will arise around unique identifiers for football players – especially outside of the top competitions – at youth and grassroots levels where players will be able to look back at their history one any data point has been matched with their FIFA Connect ID.

REFERENCES

Cavoukian, Ann (2010). Privacy by Design: The Definitive Workshop. A Foreword by Ann Cavoukian, Ph.D. *Identity in the Information Society*. 3 (2): 247–251.

European Union (2020). What is Considered Personal Data under the EU GDPR? Brussels: European Union. Retrieved from https://gdpr.eu/eu-gdpr-personal-data/

FIFA (2020a). FIFA Connect Data. Zurich: FIFA. Retrieved from https://data.fifaconnect.org

FIFA (2020b). Making Football Truly Global. Zurich: FIFA. Retrieved from https://resources.fifa.com/image/upload/making-football-truly-global-the-vision-2020-2023.pdf?cloudid=z25oyskjgrxrudiu7iym

FIFA (2020c). Regulations on the Status and Transfer of Players, June 2020 Edition. Zurich: FIFA. Retrieved from https://resources.fifa.com/image/upload/regulations-on-the-status-and-transfer-of-players-june-2020.pdf?cloudid=ixztobdwje3tn2bztqcp

3

Effects of situational variables on match statistics in elite football

— Carlos Lago[1], Miguel Ángel Gómez[2] &
Richard Pollard[3]

1: University of Vigo, Spain

2: Technical University of Madrid, Spain

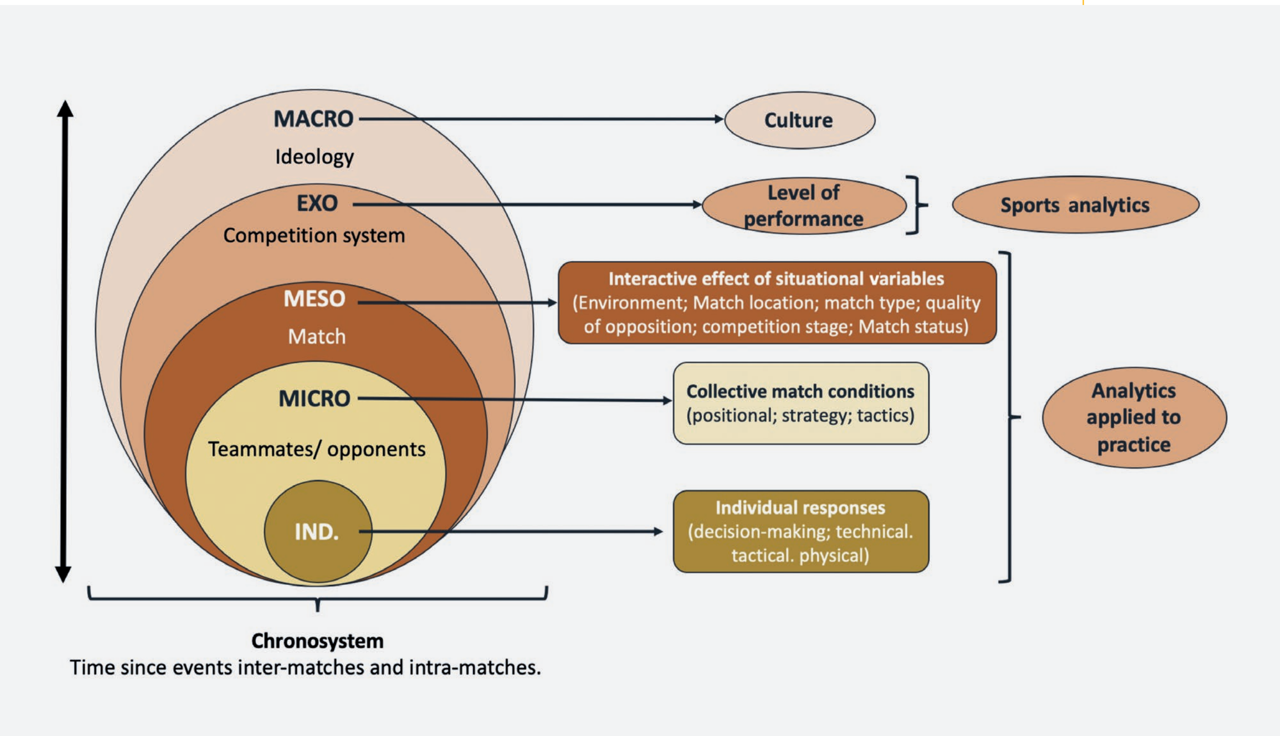
3: California Polytechmic State University, United States

INTRODUCTION

One of the most intriguing issues that has attracted attention in elite football is the impact of situational (or contextual-related) variables on teams' and players' performances (Gómez, Lago-Peñas & Pollard, 2013). This concept was initially described as 'situational variables', but some studies also considered the terms 'contextual' and 'environmental' when analysing specific conditions and scenarios during matches (Castellano, Blanco-Villaseñor & Álvarez, 2011; Lago, Casais, Domínguez & Sampaio, 2010). In particular, the available research has described how situational variables have an effect on football performance at an intra-match (i.e. match location and environmental conditions, match type, match status, or quality of opposition) and inter-match (i.e. competition stage or rounds in the league) levels (Gómez et al. 2013). Due to the complex nature of football, the analysis of situational variables cannot then be understood as the sum of isolated factors because they occur in an interacting manner during matches (Lago et al. 2010). Accordingly, this chapter will try to clarify its nature (i.e. an ecological approach) and to improve the knowledge about situational variables and their interacting effect during football matches. Thus, this approach will lead to practical information when modelling technical, tactical or physical performances according to these situational variables.

PERFORMANCE ANALYSIS IN FOOTBALL UNDER THE ECOLOGY OF HUMAN DEVELOPMENT THEORY

Sports performance analysis is a research subdiscipline of sport sciences that allows coaches, managers, performance analysts, and coaches to have a better understanding of how they perform during training and competitions (O'Donoghue, 2014). In particular, the high-order description of players' and teams' performances using performance indicators (i.e. physical and technical-tactical parameters: total distance covered or acceleration; and passing effectiveness of tackles won, respectively) allows key information from competitions to be transferred into practice and performance modelling (Rein & Memmert, 2016). The current performance approach in football can then be related to the ecology of human development theory that accounts for all the main factors that have an effect on performances, from cultural to individual levels (Bronfenbrenner, 1979). This ecological perspective is presented based on the following levels (adapted by Gómez, 2018; see figure 1): macro (the culture/idiosyncrasy of the country), exo (the competition system or the league), meso (interactive effect of situational variables during the match/es played), micro (collective match conditions according to both teams), and individual (the player's responses in each situation during the match). The importance of time (chronosystem) also exists in this ecological model according to the time during the season (inter-matches) and within each match (intra-match).



This approach allows the consideration from an ecological perspective of the analysis of sports performance in football using sport analytics when controlling for situational variables and collective/individual responses (at meso, micro and individual levels). The information obtained would then be useful to model performances according to the match and competition demands at different levels due to the ecological knowledge about how, when, where and what is measured in football.

^
Figure 1.
The application of ecology of human development to performance analysis in football (Gómez, 2018).

CLIMATIC CONDITIONS

Football is played worldwide and in highly varied climatic conditions. Many factors, such as thermoregulatory, cardiovascular and/or metabolic stress may contribute to the reductions in match performance capacity.

Playing in the heat may increase perspiration rate and peripherical vasodilation, which can result in dehydration and competition between metabolic demands and heat loss requirements. For instance, compared to thermoneutral condition (21°C), the total distance and the high-intensity running decreased respectively by 7% and 26% in elite footballers playing in the heat (43°C) (Mohr, Nybo, Grantham & Racinais, 2012). Furthermore, a decrease in percentage of total distance covered at low- to moderate-speed running has been observed when playing in 41°C compared to 35°C (Özgünen et al. 2010). This is accompanied by transient fatigue development and post-match decreases in jumping (-8.2%) and repeated-sprint (-2.6%) performances (Mohr et al. 2010). Similarly, during the last 2014 FIFA World Cup, high-intensity running activity decreased in matches performed under elevated heat stress (Nassis, Brito, Dvorak, Chalabi & Racinais, 2015) It has been suggested that such reduction may allow footballers to preserve their sprinting ability (Link & Weber, 2017) and/or to maintain successful technical skills (Nassis et al. 2015).

Exposure to altitude has detrimental effects on the human body and consequently on exercise performance (Nassis, 2013). High-speed activity and accelerations (9-25% decrease) appear to be the most susceptible to changes when matches at an altitude between 1600 and 3600 m are examined in comparison with sea level (Aughey et al. 2013; Garvican et al. 2014; Trewin, Meylan, Varley & Cronin, 2017). Total distance covered at the 2010 FIFA World Cup (3.1%) was also reduced above 1200 m compared with sea level (Nassis, 2013). There is strong evidence to believe that there is an increased home advantage when playing at high altitude. Sea-level teams playing against teams that are residents of moderate/high altitude have a low probability to win when playing away (McSharry, 2007). It has been demonstrated that altitude-induced changes in ball's flight characteristics might diminish neuromuscular coordination of high-altitude resident players when competing at sea level (Levine, Stray-Gundersen & Mehta, 2008).

MATCH LOCATION

The home advantage (HA) effect in football has appeared since the beginning of competitions in England (Pollard & Pollard, 2005), Spain, France, Italy and Portugal (Pollard & Gómez, 2009). This effect has played a key role in winning at home with values of around 60% (the proportion of points won at home from the total points won) in European leagues (Pollard & Gómez, 2009). The HA effect has been described due to the influence of some factors on the performance level of players and teams (i.e. referee bias, familiarity, territoriality, crowd effect, tactics, rule factors or psychological factors) (Pollard, 2008). This phenomenon also exists in women's football (Pollard & Gómez, 2014) and varies according to some contexts, with teams playing in isolated locations (e.g. mountains, islands or remote places) or locations with ethnical particularities (i.e. Balkans) having greater values than those teams playing in capital cities (Pollard & Gómez, 2009). In addition, the HA effect has changed over time, with higher values at the beginning of competitions and a decline of this effect during the last two decades (Pollard & Gómez, 2009). This fact may be due to the influence of professionalism, wars, the increase of recruiting foreign players (e.g. as a result of Bosman's Law), or better conditions for traveling and training (Pollard & Pollard, 2005).

The available research has identified some key performance indicators related to home teams such as ball possession, shots on target, corners or passing effectiveness (Carmichael & Thomas, 2005). Specifically, some defensive indicators (i.e. fouls or yellow cards), tactical patterns or strategies were described to differentiate home and away teams (Tenga, Holme, Ronglan & Bahr, 2010; Tucker, Mellalieu, James & Taylor, 2005). The analysis of match location is an important situational variable that affects performance at a meso level interacting with match status, competition stage, match period and quality of opposition.

COUNTRY	MEN'S HA	WOMEN'S HA
England	58.4%	49.6%
France	59.4%	53.4%
Italy	57.8%	53.6%
Germany	59.5%	53.6%
Spain	59.5%	55.9%

^
Table 1.
Home advantage in the big five leagues for the last five seasons (original data from 2014-2015 to 2018-2019 seasons).

SCORE-LINE

Football is a dynamic and complex sport where teams and players have to organise their activity according to strategic plans, principles of play and action guidelines decided upon before a match, and to adapt their tactics to the immediate requirements of an ever-changing opposition (Memmert, Lemmink & Sampaio, 2017). This means that teams have to choose a specific combination of attacking and defensive playing styles, considering their strengths and weaknesses, in order to increase their probability of success.

However, scoring or conceding a goal may change the plan and leads to important changes in the way many teams play. Performance accomplishments are a powerful source of efficacy expectations that can determine the task-related effort that has to be expended (Bandura, 1977). In football, the score-line is a measure of performance accomplishments and hence may influence at a micro level (collective match conditions: strategy, tactics) and individual level (the player's responses in each situation during the match).

At a micro level, the score-line can impact the preferred playing patterns of a team. The influence of this factor is reflected in changes in team strategies and tactics as a response to match situations. It has been demonstrated, for example, that teams tend to increase their possession when in a losing state, suggesting that they attempt to work harder to get back in the game (Lago, 2009; Lago & Martin, 2007; O'Donoghue & Robinson, 2016). The time spent in possession of the ball in different zones of the pitch (defensive third, middle third, attacking third) is influenced by the score-line: when teams are losing, possession of the ball is less in the defensive zone and more in the attacking zone than when winning or drawing. However, successful teams have been found to maintain possession whether they are winning or losing compared to unsuccessful teams (Bloomfield, Polman & O'Donoghue, 2005).

At an individual level, it has been suggested that in the hope of getting back into the game, very high-speed running increased when teams are losing a match compared to winning. For example, Lago et al. (2010) demonstrate that for every minute losing, an extra 1 m of distance was covered at sprint speed (19.1 km.-1) compared to winning. Conversely, winning increased low-speed movements (<11 Km.-1). However, it seems that this score-line effect affects players' running performance differently depending on their playing position (Bradley and Noakes, 2013; Reedwood-Brown, O'Donoghue, Robinson & Neilson, 2012; Reedwood-Brown, O'Donoghue, Nevill, Saward & Dyer, 2018). Defenders cover longer distances when losing, while attacking players show the opposite trend. These results suggest that players probably do not always use their maximal physical capacity during the match (Lago et al. 2010; Rampinini, Impellizzeri, Castagna, Coutts & Wisloff, 2009).

Coaches should control the effect of scoring or conceding a goal in the further development of the match and work with the players in these two situations. When a goal is conceded, it is about helping footballers to control their frustration that things didn't work out as planned. They must be able to stick to the plan devised for the match or modify it and adjust it to the new scenario without losing confidence in what they are doing. If they score first, they will then have to insist on keeping the same performance that has led them to this partial success during the match. Teams must have strategies prepared in order to face both scenarios. 'Psychological' goals do not exist in football; there are no goals that count more than others. It all depends on how players react to the changes in the score-line. Training will determine whether they are affected for good or bad, by helping players to control their reaction to scenarios during the match that cannot be anticipated.

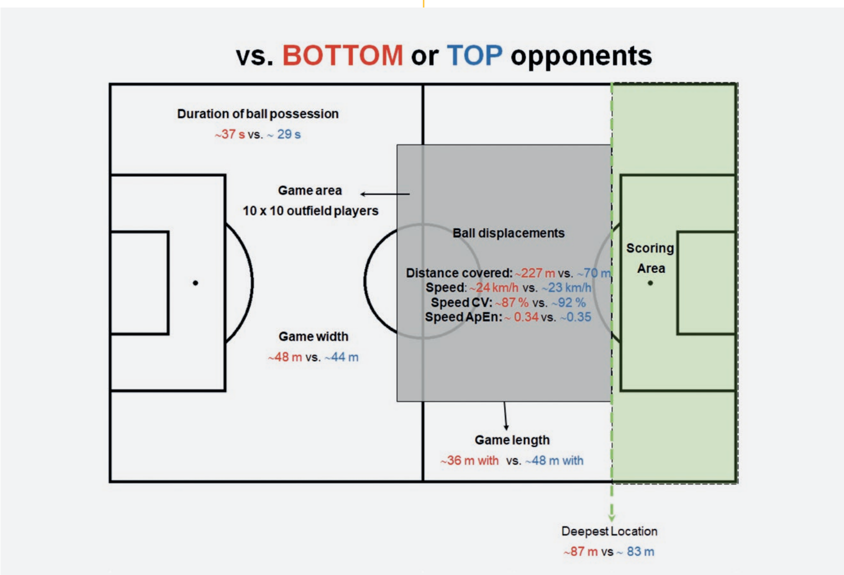
QUALITY OF OPPOSITION

Technical, tactical and physical factors may also vary as a function of the opposition's ability. Existing literature suggests that teams may alter their game strategy in relation to the standard of the opposition they are playing, but not necessarily their performance accuracy of these technical variables. For example, most of studies demonstrate that the poorer the quality of the opponent, the greater the time spent in possession by the reference team (Bradley, Lago-Peñas, Rey & Sampaio, 2014; Lago, 2009; Lago & Martin, 2007). Conversely, less successful teams are more likely to play a more defensive style against higher-ranked teams, increasing the player density within their defensive half to minimise attacking threats and opportunities (shots and crosses) (Gollan, Bellenguer & Norton, 2020). Top teams retained more possession than their opponents, suggesting they prefer to control the game by dictating play.

In a recent study, Gonçalves et al. (2019) examined a total of 1,413 ball possession sequences, obtained from 12 elite football matches from one team (the team ended the season in the top-five position). Data included the ball-possession sequences from six matches played against top opponents (TOP, the three teams classified in the first three places at the end of the season) and six matches against bottom opponents (BOTTOM, the three teams classified in the last three at the end of the season). They found that when playing against TOP opponents, there was ~38 meters game length per ~43 meters game width with 12% of coefficient of variation (%). Ball possessions lasted for ~28 seconds and tended to end at ~83m of pitch length. Against BOTTOM opponents, a decrease in the game length with an increase in game width and in the deepest location was observed in comparison with playing against TOP opponents. The duration of ball possession increased considerable (~37 seconds), and the ball-speed entropy was higher, suggesting lower levels of regularity in comparison with TOP opponents (see Figure 2).

Total running alone does not relate to winning games. However, players from more successful teams had greater ball involvement, with greater total distance and high-speed running performed whilst in possession. However, less successful teams have been reported to cover greater high- and very-high speed running distance compared to more successful teams (Lago et al. 2010; Rampinini et al. 2009). Research demonstrates that the poorer the quality of the opponent, the shorter the distance covered by the reference team in a match. Conversely, players cover more ground when their opposing team is higher in ability.

Figure 2.
A practical application example of how the quality of the opposition affects ball possession strategies. All the spatial references are real values. The dimension of the pitch is 105 × 68 meters (adapted from Gonçalves et al. 2019).
v



MATCH PERIOD

The match period is a crucial situational variable during football matches where critical moments can appear decreasing teams' and players' performances (Bar-Eli & Tracinsky, 2000; Bransen et al. 2019). Specifically, the flow of the football match can create in both teams positive and negative momentum due to match status and the period of the match (starting, middle or ending moments of the match). A psychological crisis can then decrease the team's performance during the last minutes of the match where players' decision-making may worsen (Bar-Eli, Tenembaum, & Geister, 2006). This situational variable has been analysed interacting with match location, match status and quality of opposition during football matches (Lago et al. 2010). In fact, the main results showed how teams vary their performances as the match goes on with better performances for home teams at the beginning of the match (i.e. crowd support and familiarity) and more varied performances at the end of the match based on match status and quality of opposition (Gómez et al. 2013). More recently, Gómez, Reus, Parmar & Travassos (2020) identified how football teams regain the initiative during the match according to the match status and match period, with the last periods of the match as the key to determining the winner due to increasing passing effectiveness and possession performances.

Accordingly, the importance of match period in football should be considered at the beginning, at the middle and at the end of a match when the teams' and players' performances are affected in a different way. During football matches, players and teams perform differently according to match periods with a clear impact of starting and ending moments of the match (Lago et al. 2010; Lago-Peñas et al. 2017). Therefore, performance analyses should control for this situational variable interacting with match location, match status and quality of opposition.

TYPE OF COMPETITION

Competitive professional football is played under two main tournament formats: regular seasons and knockout stages. On the one hand, regular season matches allow teams to gain or recover points as the season goes on. Research has described the balanced performance during the beginning and ending stages of the regular season (Link & Weber, 2017); On the other hand, knockout matches require the highest level of performance to avoid elimination when facing double-leg matches. Specifically, Page and Page (2007) identified a greater probability (>50%) of qualification to the next round when playing the second-leg match at home. According to this situational variable, the role of coaches when preparing matches at an exo (i.e. type of competition) and micro levels (i.e. match importance) is quite relevant to deal with consecutive matches (i.e. managing players' fatigue, substitutions or player rotation) during the regular season and with the decision making within a critical match (i.e. team's tactics and strategies) respectively.

Football is a low-scoring sport that is highly affected by this situational variable due to the high impact of a goal scored/conceded during the match. Particularly during knockout matches in which the best teams play, the quality of opposition and team ability need to be considered. Additionally, knockout matches can be played on non-typical weekdays and amid congested fixtures (i.e. UEFA Champions League or UEFA Europe League), affecting the team's performance and the home advantage (Goller & Krumer, 2020). Therefore, the interaction of these variables must be considered when accounting for teams' and players' performances (Gómez et al. 2013). Lastly, international competitions for national teams (e.g. World Cup or European Championships, or World Cup qualifications) should be considered to analyse playing styles, patterns of play and the different performances displayed during these matches (Liu, Gómez, Lago-Peñas & Sampaio, 2015; Pollard & Armatas, 2017).

INTERACTIVE EFFECTS

As was argued earlier, each situational variable has an effect on teams' and players' performances at different levels (i.e. from exo to micro levels) when studying their effect in isolation (Gómez et al. 2013). However, the main importance of these variables is the interactive effect when two or more variables affect players/teams during the match. On the one hand, this analysis requires the study of the full spectrum of situational variables to have a complete approach of sport performance (i.e. contextual-related). On the other hand, the statistical models used should account for the complex nature of those performances investigated (repeated measures for each team/player and the interaction among situational variables). The available research described how some situational variables interacted during matches. For example, the quality of opposition, match location and score-line affected ball possession and passing performances (Lago, 2007); match location and match period affected tackles or crosses (Carmichael & Thomas, 2005); and quality of opposition and competition type had an effect on crosses, ball possession or aerial duels (Castellano et al. 2011; Lago, 2009).

STUDY	LEAGUE/SAMPLE	KEY PERFORMANCE INDICATORS	STATISTICAL MODEL	SITUATIONAL VARIABLES					
				ENV	ML	SL	QO	MP	TC
Yi et al.	UEFA CL/ players	Technical/tactical	MBI	-	* †	* †	* †	* †	*
Fernández et al.	La Liga/ teams	Tactical	Two step cluster/ one way ANOVA	-	*	* †	* †	*	-
Raya-Gonzaléz et al.	Spanish U19 players	RPE	Two ways ANOVA	-	* †	-	* †	-	-
Curtis et al.	NCAAA/ players	Physical	Multilevel mixed effect model	-	-	-	*	-	* †
Lepschy et al.	Bundesliga/ close matches	Technical/tactical	One way ANOVA/ Logit regression	-	* †	-	* †	-	-
González-Rodenas et al.	English Premier League	Playing tactics	Multilevel logistic regression	-	* †	* †	* †		-
Gómez et al.	La Liga/ Close matches	Technical/tactical	Relative phase analysis	-	-	* †	-	* †	-
Gollan et al.	English Premier League	Styles of play	Clustering/ Logistic regression	-	* †	-	* †	-	-
González-Rodenas et al.	Big five leagues	Technical/tactical	Multilevel logistic regression	-	*	* †	* †		-
Oliva-Lozano et al.	La Liga/ teams	Physical	PCA/ one way ANOVA	-	* †	-	-	-	* †
Springham et al.	English championship players	Physical	MUVR	-	*	-	* †	-	* †
Granero-Gil et al.	La Liga players	Change of direction	One way ANOVA	* †	* †	-	* †	-	* †
Aquino et al.	Brazilian league	Physical/positional	Forward stepwise discriminant function	-	* †	-	* †	-	-

Note: ENV= environmental; ML= match location; SL= score-line; QO= quality of opposition; MP= match period; TC= type of competition; *= variable studied; †=statistically significant variable; MBI= magnitude based inference; PCA= principal component analysis; MUVR= Multivariate methods with unbiased variable selection.

^
Table 2.
Information from the most recent studies (from 2020) analysing situational variables in football.

In order to describe the importance of situational variables, table 2 includes the most recent studies that controlled for situational variables in elite football. The table includes the statistical model used and the main significant effect of those situational variables considered for.

From table 2, the studies reflected that no article analysed all the situational variables described in this chapter in the same approach. In fact, most of them included two to five situational variables with different significant results, where each context and type of variable may affect differently the interaction of situational variables. In addition, despite the analysis of situational variables, interaction requires multivariate analyses (e.g. multilevel analyses, clustering or logistic regression) and control for technical, tactical, positional or physical performance indicators together. Some studies neither considered those models in their analyses (e.g. ANOVA) nor a full spectrum of variables. According to these examples and available research, one of the main limitations when analysing the impact of situational variables is the use of small sample size (i.e. reduced number of matches, performance indicators and variables) and the statistical approach focused on univariate analyses. Therefore, further research of football analytics needs to account for those limitations increasing the meaning of findings and the impact into practice of the most relevant results.

PRACTICAL APPLICATIONS

From a practical application perspective, two aspects need to be considered: on the one hand, the statistical models should control: (i) for the team/players' repeated measures performance (random effect); (ii) the interaction of situational variables as fixed factors or covariates; (iii) the use of large sample size of both observations and performance indicators; and (iii) the combination of multivariate models that allow to classify (e.g. clustering), differentiate (e.g. discriminant analyses) and predict (e.g. regression) performances. On the other hand, the analysis of the context of the most relevant variables to study at the different levels from exo (e.g. league), meso (interactive effect of situational variables), micro (collective match conditions), and individual (the player's responses during the match) considering the temporal series of matches.

CONCLUDING REMARKS

Research examining situational variables such as match status (win, draw, lose) and location (home, away) level of opposition (top, middle, bottom), match half, type of competition (regular season, knockout stage), and environmental conditions (altitude, temperature) demonstrates these have an impact of the physical, technical and tactical profiles of players and teams. It seems likely no single study can comprehensively measure and control for all extraneous influences. Effective evaluation of football performance at a behavioural level needs to account for the potential interactions between variables. This should not deter researchers, however, from exploring this area with the possibility of at least establishing a hierarchy with regards to these factors (Paul, Bradley & Nassis, 2015). When controlling for situational variables, performance analysis should be carried out considering an ecological perspective on the following levels: macro (the culture/idiosyncrasy of the country), exo (the competition system or the league), meso (interactive effect of situational variables during the match/es played), micro (collective match conditions according to both teams), and individual (the player's responses in each situation during the match). Coaches and researchers need to consider all these factors before making inferences based on data supplied by match-analysis companies. In addition, probably more fine-grained approaches (e.g. accounting for repeated measures and multivariate models) to football analysis should adopted by considering a single team's performances.

REFERENCES

Aquino, R., Carling, C., Palucci, L.H., Martins, G., Jabor, G., Machado, J., Santiago, P., Garganta, J., & Puggina, E. (2020). Influence of Situational Variables, Team Formation, and Playing Position on Match Running Performance and Social Network Analysis in Brazilian Professional Soccer Players. *Journal of Strength and Conditioning Research*, 34(3): 808-817.

Aughey, R.J., Hammond, K., Varley, M.C., Schmidt, W.F., Bourdon, P.C., Buchheit, M. . . . Gore, C.J. (2013). Soccer Activity Profile of Altitude Versus Sea-Level Natives During Acclimatisation to 3600 m (ISA3600). *British Journal of Sports Medicine*, 47, 107–113

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bar-Eli,M., Tenenbaum, G., and Geister, S. (2006). Consequences of Players' Dismissal In Professional Soccer: A Crisis-Related Analysis of Group-Size Effects. Journal of Sports Sciences, 24(10), 1083-1094

Bloomfield, J.R., Polman, R.C.J., & O'Donoghue, P.G. (2005b). Effects of Score-line on Team Strategies in FA Premier League *Soccer. Journal of Sports Sciences*, 23, 192-193.

Bradley P.S., Lago-Peñas C., Rey E., & Sampaio, J. (2014) The Influence Of Situational Variables on Ball Possession in the English Premier League. *Journal of Sports Sciences*, 32, 1867-1873.

Bradley, P.S., & Noakes, T.D. (2013). Match Running Performance Fluctuations In Elite Soccer: Indicative of Fatigue, Pacing or Situational Influences? *Journal of Sports Sciences*, 31, 1627-1638.

Bransen, L., Robberechts, P., Van Haaren, J., & Davis, J. (2019). Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. In *Proceedings of the 13th MIT Sloan Sports Analytics Conference* (pp. 1-25). MIT SLOAN; http://www.sloansportsconference.com/wp-content/uploads/2019/02/Choke-or-Shine-Quantifying-Soccer-Players-Abilities-to-Perform-Under-Mental-Pressure.pdf

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge, Massachusetts and London: Harvard University Press.

Carmichael, F., & Thomas, D. (2005). Home Field Effect and Team Performance: Evidence from English Premiership football. *Journal of Sports Economics*, 6(3), 264–281.

Castellano, J., Blanco-Villaseñor, A., & Alvarez, D. (2011). Contextual Variables and Time-Motion Analysis in Soccer. *International Journal of Sports Medicine*, 32(06), 415-421.

Curtis, R. M., Huggins, R. A., Benjamin, C. L., Sekiguchi, Y., Adams, W. M., Arent, S. M., ... & Casa, D. J. (2020). Contextual Factors Influencing External and Internal Training Loads in Collegiate Men's Soccer. *The Journal of Strength & Conditioning Research*, 34(2), 374-381.

Fernandez-Navarro, J., Ruiz-Ruiz, C., Zubillaga, A., & Fradua, L. (2019). Tactical Variables Related to Gaining the Ball in Advanced Zones of the Soccer Pitch: Analysis of Differences Among Elite Teams and the Effect of Contextual Variables. *Frontiers in Psychology*, 10.

Garvican, L., Hammond, K., Varley, M., Gore, C., Billaut, F., & Aughey, R. (2014). Lower Running Performance and Exacerbated Fatigue in Soccer Played at 1600 m. *International Journal of Sports Physiology and Performance*, 9, 397–404.

Goller, D., & Krumer, A. (2020). Let's Meet as Usual: Do Games Played on Non-Frequent Days Differ? Evidence From Top European Soccer Leagues. *European Journal of Operational Research*, 16(2), 740-754.

Gollan, S., Bellenger, C., & Norton. K. (2020) Contextual Factors Impact Styles of Play in the English Premier League. *Journal of Sports Science and Medicine*, 19, 78 - 83.

Gómez, M.A. (2018). Science Into Practice in Performance Analysis in Sport. World Congress of Performance Analysis in Sport, Opatija, Croatia.

Gómez, M. Á., Lago-Peñas, C., & Pollard, R. (2013). Situational Variables. In T McGarry, P. O'Donoghue & J Sampaio (Eds.). *Routledge Handbook Of Sports Performance Analysis* (pp. 277-287). Routledge.

Gómez, M. A., Reus, M., Parmar, N., & Travassos, B. (2020). Exploring Elite Soccer Teams' Performances During Different Match-Status Periods Of Close Matches' Comebacks. *Chaos, Solitons & Fractals*, 132, 109566.

Gonçalves, B., Coutinho, D., Exel, J., Travassos, B., Lago, C. & Sampaio, J. (2019). Extracting Spatial-Temporal Features That Describe a Team Match Demands When Considering the Effects of the Quality of Opposition In Elite Football. PLoS ONE 14(8): e0221368.

González-Ródenas, J., Aranda, R., & Aranda-Malaves, R. (2020). The Effect of Contextual Variables on the Attacking Style of Play in Professional Soccer. *Journal of Human Sport and Exercise*, in press.

Gónzalez-Ródenas, Mitrotasios, M., Aranda. R., & Armatas, V. (2020). Combined Effects of Tactical and Contextual Factors on Shooting Effectiveness in European Professional Soccer. *International Journal of Performance Analysis in Sport*, 20(2), 280-293.

Granero-Gil, P., Gómez-Carmona, C.D., Bastida-Castillo, A., Rojas-Valverde, D., de la Cruz, E., & Pino-Ortega, J. (2020) Influence of Playing Position and Laterality in Centripetal Force and Changes of Direction in Elite Soccer Players. *PLoS ONE*, 15(4): e0232123.

Lago, C. (2009). The Influence of Match Location, Quality of Opposition, and Match Status on Possession Strategies in Professional Association Football. *Journal of Sports Sciences*, 27, 1463-1469.

Lago, C., Casais, L., Dominguez, E., & Sampaio, J. (2010). The Effects of Situational Variables on Distance Covered at Various Speeds in Elite Soccer. European *Journal of Sport Science*, 10, 103-109.

Lago-Peñas, C., & Dellal, A. (2010). Ball Possession Strategies in Elite Soccer According to the Evolution of the Match-Score: The Influence of Situational Variables. *Journal of Human Kinetics*, 25, 93-100.

Lago, C., & Martin, R. (2007). Determinants of Possession of the Ball in Soccer. *Journal of Sports Sciences*, 125, 969-974.

Levine, B.D., Stray-Gundersen, J., & Mehta, R.D. (2008). Effect of Altitude on Football Performance. *Scandinavian Journal of Medicine & Science in Sports*, 18, 76–84.

Lepschy, H., WÄsche, H., & Woll, A. (2020). Success Factors in Football: An Analysis of the German Bundesliga. *International Journal of Performance Analysis in Sport*, 20(2), 150-164.

Link, D., & Weber, H. (2017). Effect of Ambient Temperature on Pacing in Soccer Depends on Skill Level. *Journal of Strength and Conditioning Research*, 31(7), 1766–1770.

Liu, H., Gomez, M. Á., Lago-Peñas, C., & Sampaio, J. (2015). Match Statistics Related to Winning in the Group Stage of 2014 Brazil FIFA World Cup. *Journal of Sports Sciences*, 33(12), 1205-1213.

Memmert, D., Lemmink, K. A., & Sampaio, J. (2017). Current Approaches To Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47, 1–10.

McSharry, P.E. (2007). Altitude and Athletic Performance: Statistical Analysis Using Football Results. BMJ 335, 1278–1281.

Mohr, M., Mujika, I., Santisteban, J., Randers, M.B., Bischoff, R., Solano, R., Hewitt, A., Zubillaga, A., Peltola, E., & Krustrup, P. (2010). Examination of Fatigue Development in Elite Soccer in a Hot Environment – A Multiexperimental Approach. *Scandinavian Journal of Medicine & Science in Sports*, 20 (Suppl. 3), 125– 132.

Mohr, M., Nybo, L., Grantham., J., & Racinais, S. (2012). Physiological Responses and Physical Performance During Football in the Heat, *PLoS One*. 7, e39202.

Nassis, G.P. (2013) Effect of Altitude on Football Performance: Analysis of the 2010 FIFA World Cup Data. *Journal of Strength and Conditioning Research*, 27, 703-707.

Nassis, G.P., Brito, J., Dvorak, J., Chalabi, H., & Racinais, S, (2015). The Association of Environmental Heat Stress with Performance: Analysis of the 2014 FIFA World Cup Brazil. *British Journal of Sports Medicine*. 49, 609–613.

O'Donoghue, P. (2014). *An introduction to performance analysis of sport*. London: Routledge.

O'Donoghue P, Robinson G. (2016). Score-line Effect on Work-rate in English FA Premier League soccer. *International Journal of Performance Analysis in Sport*, 16, 910-923.

Özgünen, K.T., Kurdak, S.S., Maughan, R.J., Zeren, C., Korkmaz, S., Yazici, Z., Ersöz, G., Shirreffs, S.M., Binnet, M.S., & Dvorak, J. (2010). Effects of Hot Environmental Conditions on Physical Activity Patterns and Temperature Response of Football Players. *Scandinavian Journal of Medicine & Science in Sports*, 20, 140–147.

Oliva-Lozano, J.M., Rojas-Valverde, D., Gómez-Carmona, C.D., Fortes, V., & Pino-Ortega, J. (2020). Impact of Contextual Variables on the Representative External Load Profile of Spanish Professional Soccer Match-Play: A Full-Season Study. *European Journal of Sport Science*, in press.

Page, L., & Page, K. (2007). The Second Leg Home Advantage: Evidence from European Football Cup Competitions. *Journal of Sports Sciences*, 25(14), 1547–1556.

Paul, D.J., Bradley, P.S., & Nassis, G. (2015). Factors Influencing Running Performance of Elite Soccer Players: Shedding Some Light on the Complexity. *International Journal of Sports Physiology and Performance*, 10, 516-519.

Pollard, R. (2008). Home Advantage in Football: A Current Review of an Unsolved Puzzle. *The Open Sports Sciences Journal*, 1, 12-14.

Pollard, R., & Gómez, M. A. (2009). Home Advantage in Football in South-West Europe: Long-Term Trends, Regional Variation, and Team Differences. *European Journal of Sports Sciences*, 9, 341-352.

Pollard,R., & Gómez, M.A. (2014). Components of Home Advantage in 157 National Soccer Leagues Worldwide, *International Journal of Sport and Exercise Psychology* 12(3), 218-233.

Pollard, R., & Pollard, G. (2005). Long Term Trends in Home Advantage in Professional Team Sports in North America and England (1876 – 2003). *Journal of Sports Sciences*, 23, 337-350.

Pollard, R., & Armatas, V. (2017). Factors Affecting Home Advantage in Football World Cup Qualification. *International Journal of Performance Analysis in Sport*, 17(1-2), 121-135.

Rampinini, E., Impellizzeri, F.M., Castagna, C., Coutts, A.J. & Wisloff, U. (2009) Technical Performance During Soccer Matches Of The Italian Serie A League: Effect of Fatigue and Competitive Level. *Journal of Science and Medicine in Sport* 12(1), 227-233.

Raya-González, J., Castillo, D., Yanci, J., & Los Arcos, A. (2020). Assessing the Perceived Exertion in Elite Soccer Players During Official Matches According to Situational Factors. *International Journal of Environmental Research and Public Health*, 17(2), 410.

Redwood-Brown, A., O'Donoghue, P., Robinson, G., & Neilson, P. (2012). The Effect Of Score-Line on Work-Rate in English FA Premier League Soccer. *International Journal of Performance Analysis in Sport*, 12, 258-271.

Reedwood-Brown, A.J., O'Donoghue, P.G., Nevill, A.M., Saward, C., Dyer, N., & Sunderland, C. (2018). Effects of Situational Variables on the Physical Profiles of Elite Soccer Players in Different Score-Lines States. *Scandinavian Journal of Medicine & Science in Sports*, 28, 2515-2526.

Rein, R., & Memmert, D. (2016). Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science. *SpringerPlus*, 5(1), 1-13.

Springham, M., Williams, S., Waldron, M., Strudwick, A.J., Mclellan, C., & Newton, R.U. (2020). Prior Workload has Moderate Effects on High-Intensity Match Performance in Elite-Level Professional Football Players when Controlling for Situational and Contextual Variables. *Journal of Sports Sciences*, in press.

Tenga, A., Holme, I., Ronglan, L.T., & Bahr, R. (2010). Effects of Match Location on Playing Tactics for Goal Scoring in Norwegian Professional Soccer. *Journal of Sport Behavior*, 33(1), 89–108.

Trewin, J., Meylan, C., Varley, M.C., & Cronin, J. (2017). The Influence of Situational and Environmental Factors on Match-Running in Soccer: A Systematic Review, *Science and Medicine in Football*, 1:2, 183-194.

Tucker,W., Mellalieu, S.D., James, N., & Taylor, J.B. (2005). Game Location Effects in Professional Soccer. A Case Study. *Internatinal Journal of Peformance Analysis in Sports*, 5, 23-35.

Yi, Q., Gómez, M.A., Liu, H., Gao, B., Wunderlich. F, & Memmert D (2020). Situational and Positional Effects on the Technical Variation of Players in the UEFA Champions League, *Frontiers in Psychology*, 11:1201.

How does context affect player performance in football?

— Lotte Bransen[1], Pieter Robberechts[2], Jesse Davis[2],
Tom Decroos & Jan Van Haaren[1,2]

1: SciSports
2: KU Leuven

INTRODUCTION

Football professionals are faced with many challenges. In order to make the best possible decision in each situation, they need to consider the contexts that may impact their decisions. Although technological advances are increasingly providing insights from data, many widely used performance metrics still struggle to capture important contextual information. As a result, the focus in the football analytics community has recently shifted to developing methods that better capture the circumstances in which players perform their actions. This chapter overviews recent approaches that quantify player performance and to what extent these approaches account for relevant technical, tactical and mental aspects that affect player performance.

Football professionals and fans are faced with a multitude of important questions on a daily basis. Managers and coaches are tasked with fielding the best possible line-up in their teams' matches, scouts and recruitment analysts are concerned with identifying potential transfer targets that suit their teams' playing styles, player representatives and agents aim to sign emerging youngsters at the start of their careers, and football fans aim to compose the best-performing fantasy football teams. For all these tasks, it is crucial to consider the relevant contexts in order to make the best possible decision. For instance, in scouting, the quality of the opposition that potential transfer targets played against is one relevant context. Hence, a scout must evaluate their performances in this light and analyse how these performances would relate to the quality of the opposition they would face at their new club.

To satisfy the desire of practitioners to better understand the game, the recent focus in the football analytics community has been on better capturing the context in which players perform their actions. The increasing granularity of football data has enabled metrics to account for technical, tactical and mental aspects that impact

the performances of football players. Examples include the consideration of players' finishing abilities (Gregory, 2017; Kwiatkowski, 2017), teams' playing styles (Decroos et al. 2018), and players' mental resilience in performance metrics (Bransen et al. 2019).

This chapter overviews how context affects the performances of football players by addressing the following four questions in turn in the following sections:

- How to quantify the impact of a single action?
- How do technical aspects affect a player's performance?
- How do tactical aspects affect a player's performance?
- How do mental aspects affect a player's performance?

HOW TO QUANTIFY THE IMPACT OF A SINGLE ACTION?

Due to the low-scoring nature of football, the impact of the many actions that a player performs in a match is hard to quantify in isolation. As a result, the focus in the football analytics community has gradually shifted from quantifying each action's impact on the goals that were scored (e.g. counting assists or pre-assists) to quantifying each action's impact on increasing or decreasing the likelihood of scoring a goal from a particular game situation (e.g. computing xG values for shots). By doing so, the performance of a player can be assessed from a much larger, more representative, set of actions.

To adequately measure an action's impact, it is crucial to reason about the context in which the action was performed. For example, a foul committed in the penalty box increases the odds of conceding a goal in most situations, but avoids an almost certain goal in situations where the goalkeeper is out of position. Inspired by the richness of present-day football data, the field of football analytics has shifted in recent years towards devising performance metrics that incorporate the relevant contextual information to quantify player performance. The remainder of this section outlines the most relevant approaches to quantify the impact of shots, passes and other on-the-ball actions as well as off-the-ball actions.

SHOTS

Shot-based performance metrics have quickly evolved from context-free metrics such as the shot conversion rate (i.e. the fraction of a player's shots that finds the net) to context-heavy metrics such as xG values. Since finishing a close-range shot with no defenders around is typically easier than scoring from a long-range shot with a crowded penalty box, a player usually deserves more credit for scoring the long-range shot than for scoring the easy shot from close to the goal. Traditionally, the impact of a shot is quantified by contrasting its xG value (i.e. the likelihood of a shot yielding a goal) with its actual outcome (i.e. goal or no goal). As a result, a player is positively rewarded for scoring a difficult long-range shot and negatively rewarded for missing a sitter.

xG models have evolved over time from considering little contextual information, such as the location of the shot (Green, 2012), to more sophisticated models that account for the locations of opponents and the height of the ball at the time of the shot (Knutson, 2020). The early xG models mostly include characteristics of the shot, such as the distance to the goal, the shot angle and whether the shot was headed or not (Green, 2012; Pollard

& Reep, 1997). These models distinguish between short- and long-distance shots and recognise that shots with the foot are easier to finish than shots with the head. However, they do ignore important contextual information, such as the build-up to the shot. For instance, scoring from a fast counter-attack is often easier than scoring from a slow build-up when the defense has more time to get organised. Therefore, most early xG models comprise an ensemble of regression models that each address particular types of situations, such as footed open-play shots, headed open-play shots, and direct free kicks (Caley, 2015; Ijtsma, 2015). The introduction of more granular match event data and spatio-temporal tracking data has eventually led to models that can identify, for instance, shots taken under pressure of defenders, shots where the goalkeeper is out of position and shots that require difficult body movements from the goalkeeper to reach the ball (Knutson, 2018, 2020; Lucey et al. 2015).

Although xG models are primarily used to evaluate the performances of attackers, they are also increasingly used to evaluate the performances of goalkeepers. While xG values provide insights into how many goals an attacker should have scored, expected-saves values (xS values) provide insights into how many on-target shots a goalkeeper should have saved. Attackers are typically positively rewarded for scoring more goals than expected, whereas goalkeepers are typically positively rewarded for conceding fewer goals than expected. However, two important differences exist between evaluating attackers and goalkeepers. First, unlike attackers, goalkeepers are only evaluated based on the shots that hit the target and could possibly be saved. Second, while xS models are similar in spirit to xG models, they usually also take relevant post-shot information into account, such as the direction of the shot (Lawrence, 2015; Ruiz et al. 2017; Trainor, 2014).

PASSES AND OTHER ACTIONS

Since shots constitute less than 2% of the on-the-ball actions in a match and primarily involve attacking players and goalkeepers, assessing the performances of defenders and midfielders remains hard using the aforementioned xG and xS models. Other types of on-the-ball actions happen much more often, with passes being the most frequent type of action (Decroos et al. 2019). As a result, quantifying the impact of other types of on-the-ball actions – including passes, dribbles, interceptions and tackles, among others – has gained more attention in recent years. Widespread count-based metrics are limited as they ignore important contextual information. For example, metrics such as passing accuracy and number of completed passes ignore the circumstances in which passes are executed. For example, a lateral pass from one central defender to the other is different from a perfect through ball from a midfielder that puts the striker in a one-on-one with the goalkeeper. While both passes would be counted as accurate, the latter pass has a considerably higher positive impact on the team's scoring chances than the former pass.

Analogous to shots, the general approach to quantifying the quality of an arbitrary on-the-ball action is to measure to what extent the action increases or decreases the likelihood of scoring a goal in the near future. The commonality among most of

these approaches is that they value the game states right before and after the action. Intuitively, the game state can be thought of as the current situation in the match, which may include the locations of the players and the ball, the scoreline and the time remaining in the match. Since an action transitions a match from one game state to another, its value corresponds to the difference in value between the game states before and after the action. Hence, an action that transitions the match to a more favorable game state is positively rewarded, whereas an action that transitions the match to a less favorable game state is negatively rewarded.

Different approaches have been proposed to value game states with the key difference being the amount of contextual information that each approach takes into account. Some approaches solely consider the sequence that an action belongs to Lawrence, 2018; Decroos et al. (2017), whereas other approaches also consider the location of the ball (Singh, 2019; Gyarmati & Stanojevic, 2016). However, these approaches cannot distinguish between, for example, a forward pass on the own half after regaining possession that initiates a counter-attack on one hand, and a dribble in the build-up phase without any other players around the ball on the other hand. Therefore, several approaches that leverage more contextual information to determine the game state value have been proposed. These approaches can differentiate between different match situations (e.g. penalty, corner, free kick – Rudd, 2011; Yam, 2019), type of action (Kothari, 2020) and include the trajectory and speed of play to describe the game state (Michalczyk, 2018; American Soccer Analysis, 2020; Bransen, 2017; Bransen & Van Haaren, 2018; Decroos et al. 2019; Liu et al. 2020; Mackay, 2017, 2019). These approaches can better differentiate between the aforementioned actions, but still have difficulties capturing contextual information such as pressure on the ball and players possibly blocking the pass lane.

To account for this contextual information, Power et al. (2017), introduce an approach to measure the impact of single passes using spatio-temporal tracking data. They account for factors such as the pressure on the player in possession of the ball, and whether any players are between the passing player and the intended recipient. Moreover, they consider tactical concepts, such as whether the pass is part of a counter-attack. Similarly Goes et al. (2018), use spatio-temporal tracking data to value passes. They introduce a defensive disruptiveness score to value the effectiveness of passes in disrupting the opponent's defensive organisation. Fernández et al's EPV metric (2019) and Link et al's Action Value (2016) use spatio-temporal tracking data to value actions besides passes. In their work, Fernández et al state that "a critical aspect to properly evaluate football situations is to have a clear understanding of the ongoing context". Both approaches account for the locations of all players to value on-the-ball actions.

The aforementioned approaches have been shown to identify players who contribute a lot of value to their teams and at the same time are also considered to be impactful players by the public and on the transfer market. Decroos et al's VAEP ratings identify Marcus Rashford, Trent Alexander-Arnold, Mason Mount, Kylian Mbappé and Frenkie de Jong as emerging talents in the 2017/2018 season (Decroos et al. 2019), and, Similarly, Gyarmati and Stanojevic (2016) identify Lucas Vázquez, Antoine Griezmann and Lionel Messi, among others, as the top-ranked attackers in terms of pass contribution in the 2015/2016 Spanish La Liga season. In addition, Lawrence's xG Chain metric identifies Lionel Messi, Mohamed Salah and Arjen Robben as top contributors in the 2016/2017 season (Lawrence, 2018) and Mackay's 'xG added' model identifies Eden Hazard, Alexis Sánchez and David Silva as the top contributors in the 2016/2017 English Premier League season (Mackay, 2017).

OFF-THE-BALL ACTIONS

The strong focus on on-the-ball actions is mainly driven by the widely available match event data's restriction to only registering on-the-ball actions. However, the availability of spatio-temporal tracking data describing all movements of the players and the ball has enabled the additional consideration of the impact of players' actions without the ball. For example, a player might create space for a teammate with a run or a player might mark their opponent close in a situation where the team is out of possession. However, little work has been done in this area due to the limited availability of spatio-temporal tracking data and the complexity of the problem. It is hard to divide the credit among all 22 players on the pitch when multiple players are responsible for moving the game from one game state to the other.

The EPV framework (Fernández et al. (2019) mentioned in the previous subsection makes it possible to rate players' off-the-ball impact. When looking at 'EPV added' one can credit players for the off-the-ball movements they make, even when not receiving the ball.

Dick & Brefeld (2019) use deep reinforcement learning techniques to rate player positioning. Robberechts (2019) introduces the VPEP (Valuing Pressing by Estimating Probabilities) metric that values players' pressure actions when they are not in possession of the ball.

Little work has been done in measuring the defensive impact of player movement without the ball (Llana et al. 2020; SciSports, 2020). This is partly due to the limited availability of tracking data, but mostly due to the complexity of determining who deserves credit for preventing an opponent from performing well. Consider the example where an opponent striker scores from a cross. Who is to blame? The defender who let the opponent give the cross, the two central defenders who let the striker tip in the cross or the goalkeeper who failed to prevent the ball from finding the net?

To objectively measure players' performances in matches, more and more contextual information could be incorporated in the metrics. However, to understand how and in what situations a player performs in a certain way, we describe three different types of context that could affect players' performances: technical, tactical and mental aspects. The next three sections discuss each of these types of context and their impact on player performance.


Rakuten
31
Estrella Galicia 0,0

HOW DO TECHNICAL ASPECTS AFFECT A PLAYER'S PERFORMANCE?

One crucial contextual piece of information to consider is which player performed a certain action. Each player has different technical abilities, and this will affect his or her ability to execute a particular action. For example, Kevin De Bruyne is an outstanding passer and is able to attempt and complete passes that others cannot. While not purely technical skill, players such as Eden Hazard, James Milner and Bruno Fernandes are known as excellent penalty takers who convert a higher proportion of their penalties than the typical player.

Considering the context of a player's technical skills has been most explored and hotly debated when considering finishing skills. That is, if two players could repeatedly take the same shot (i.e. the shot occurs in an identical match situation), would one player convert more chances? Intuitively, this would reflect a player's technical ability to hit the ball so that it is (a) on target and (b) placed in a location that optimises the chance of scoring (i.e. the ball is not saved by the keeper). Unfortunately, there are two big challenges related to investigating finishing. First, it is impossible to collect such data as few, if any, shots occur in identical circumstances. The standard proxy is to use an xG model, which accounts for some aspects of a shot's context. Second, very few individual players take a large number of shots, even when aggregating across different seasons. One way to identify good finishers is to look for players who score more goals than predicted by an xG model. Kwiatkowski (2017) approached the problem by training an xG model that included a set of variables representing which player took the shot. The intuition is that the weight associated with a player's variable captures his finishing skill. This is reflected in the model by the fact that if two players take an identical shot, the model will assign a higher xG to the one that is a better finisher. He found a small but noticeable effect, with players like Antoine Griezmann, Lionel Messi and Luis Suárez topping the list.

The inverse of the question can be asked about goalkeepers: How does a keeper's technical skill affect the chance that a shot will result in a goal? Here, technical skills could include executing a dive or the ability to time a jump. This is a challenging question to address for exactly the same reasons that arise when trying to assess finishing skill. First, keepers face different shots. Second, keepers face relatively few on-target shots in a season. Power et al. (2017) addressed this problem by training an expected saves (xS) model. The model captures the context of the scoring attempt by using the location of the shot and keeper at the start of the attempt as well as the score, time and location of the ball when it crosses the line or is saved. Finally, the model is augmented by considering properties of the keeper such as save percentages for various types of shots. Such a model addresses counterfactual questions, such as "what is the expected number of goals a team would have conceded with a different goalkeeper?" For example, it makes it possible to estimate how many goals Real Madrid would have conceded in the 2019-2020 season if they had not had Thibaut Courtois in goal, which in turn makes it possible to assess whether his absence would have affected their chance of winning the title.

More recently, Bransen et al. (2019) proposed an execution rating metric that attempts to answer the question: How well did the player perform the chosen action? This extends the idea of investigating finishing ability to other on-the-ball actions. Intuitively, their metric attempts to reward players who successfully perform difficult actions, such as completing a through ball or dribbling through pressure. 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, they train a model to predict the probability that an action will be successful (e.g. did the cross reach a team mate or did the player retain possession after a take-on) based on the context under which the action was performed. Here, the context is captured by features such as the player's current location on the pitch, the body part used to execute the action, and information about the previous action. The execution rating is then simply the difference between the result of the action (i.e. success – such as completing a pass or connecting on a shot – or failure, such as having a pass intercepted or missing the shot) and the predicted probability that the action would be successful. As an example, the model assigns a high execution rating to Lionel Messi's perfect free kick in the 2018/2019 Champions League semi-final against Liverpool. Table 1 shows the field players on teams in the knock-out rounds of the 2018/2019 Champions League that had the highest average pass execution ratings. The table is dominated by well-known deep-lying playmakers Toni Kroos, Axel Witsel, Steven N'Zonzi, Sergio Busquets and Rodri. These players are known for dictating the game from the midfield and their strong passing skills. The list is completed by offensive wing backs Noussair Mazraoui, Alex Sandro and Sergi Roberto; striker Karim Benzema, and centre back Matija Nastasić.

	PLAYER	CLUB
1	Toni Kroos	Real Madrid
2	Axel Witsel	Borussia Dortmund
3	Noussair Mazraoui	AFC Ajax
4	Karim Benzema	Real Madrid
5	Alex Sandro	Juventus FC
6	Steven N'Zonzi	AS Roma
7	Sergio Busquets	FC Barcelona
8	Sergi Roberto	FC Barcelona
9	Rodri	Atlético Madrid
10	Matija Nastasić	FC Schalke 04

^
Table 1.
The top 10 field players in terms of the highest average pass execution rating that played on teams in the knock-out rounds of the 2018/2019 Champions League.

HOW DO TACTICAL ASPECTS AFFECT A PLAYER'S PERFORMANCE?

Most performance metrics aim to assess individual players, although their performances are often strongly influenced by tactical concepts. Ideally, performance metrics would account for the tactical instructions that the players received. For instance, a player who is instructed to play a direct type of football should not necessarily be punished for sending risky long balls into their opponents' penalty box. However, the teams' game plans are usually not publicly known, which has inspired football analytics researchers to develop data-driven methods to automatically detect tactical concepts and playing styles.

Since detecting tactical concepts in a data-driven fashion is extremely challenging due to the fluid nature of football, the contextualisation of player performance with respect to tactics has received little attention to date. Instead, the focus has mostly been on detecting frequently occurring patterns of play in the data and measuring players' involvement in those patterns (Bekkers & Dabadghao, 2017; Decroos et al, 2018; Perdomo Meza, 2017). Some of those approaches do account for high-level relevant context (e.g. whether a team plays at home or on the road; whether a team plays a stronger or weaker opponent). Extracting recurring patterns from football data is challenging on its own and would be further complicated by accounting for player-level context (e.g. the players' playing styles).

Despite the challenges, a number of approaches addressing several tactics-related tasks have appeared to date, including techniques to automatically detect formations (Bialkowski, Lucey, Carr, Yue, & Matthews, 2014; Bialkowski, Lucey, Carr, Yue, Sridharan, et al, 2014a) as well as playing styles of players and teams (Aalbers & Van Haaren, 2018; Bialkowski, Lucey, Carr, Yue, Sridharan, et al. 2014b; Decroos et al, 2020; Decroos & Davis, 2019). Bialkowski et al. (2014) introduced an approach to automatically detect and visualise team formations. They show that teams tend to play higher up the field at home than on the road, giving merit to the common saying that teams try to win at home and draw away. Decroos et al. (2020) introduced an approach to capture the playing style of a team or player within a single match. They showed how Liverpool adopts a different playing style against Manchester City than against lesser opposition.

In addition, a number of approaches have appeared to quantify team balance by capturing each player's performance with respect to their teammates. Beal et al. (2020) presented a method that values the contributions of players to their team and automatically forms the optimal team by leveraging frequently appearing patterns in the team's passing network. Bransen and Van Haaren (2020) presented a method to quantify the chemistry between a pair of players by measuring the impact of their joint actions on their team's chances of scoring and preventing goals. Furthermore, they introduce a method that predicts the mutual chemistry between a pair of players who has never played together before. Inspired by Beal et al. (2020), they also presented a method to automatically assemble a starting line-up that maximises the mutual chemistry between pairs of players for a given squad of players. They found that Mesut Özil's performances for Arsenal rapidly declined after Alexis Sánchez' departure to rivals Manchester United in the winter of 2018. The pair had gelled particularly well in the preceding seasons and had a considerable joint impact on Arsenal's performances at the time.

HOW DO MENTAL ASPECTS AFFECT A PLAYER'S PERFORMANCE?

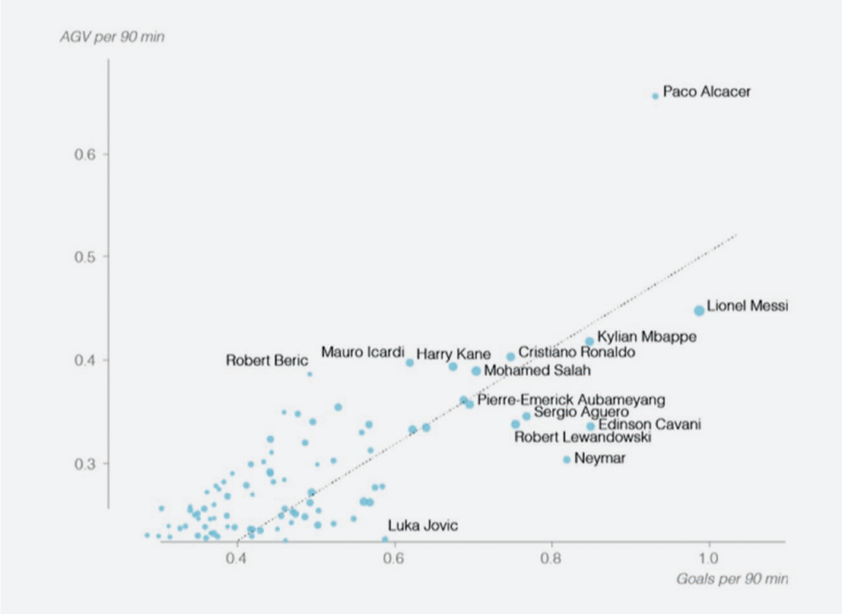
While most existing football performance metrics focus on players' physical, technical and tactical performances, they typically ignore the mental aspect. Yet, there are various mental factors that can affect a player's performance. These factors can range from mental fatigue due to the quick succession of games (Russell et al. 2019), to mental pressure in crucial game situations (Schweickle et al. 2020) or frustration when things do not go according to plan (Cowden & Worthington, 2019). In short, anything that affects the cognitive thoughts of a player would come under the mental factor. To gain a better insight in the performance of football players, it is important to recognise and develop an understanding of how these factors are presented and how they affect performance.

The vast majority of research on these mental aspects appears in cognitive sciences literature (Hill et al. 2010). This research is mostly experimental and induces cognitive stress through artificial manipulation, distraction and self-focus. For example, Williams et al. (2002) showed that anxiety is thought to affect detrimentally sporting tasks which rely heavily on decision-making by measuring variables such as reaction time, visual search data and arm kinematics in an artificial table tennis setting. The remaining body of research in this domain is qualitative and predominantly involves interviewing athletes.

In a sports analytics setting, it is much harder to quantify the mental context. While teams are able to enquire about the mental well-being of their own players, they are certainly not able to do this for opponents and future transfer prospects. Therefore, the focus is entirely on the effect of mental pressure, which is most straightforward to measure by means of proxies. With these, two groups of factors that affect the pressure level can be distinguished (Bransen et al. 2019):

- The **in-game context** or the events in the game itself. 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 score differential is high as a goal would only have a small impact on the expected outcome.

- The **pre-game context** or the context surrounding the game. 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.



^
Figure 1.
The relation between goals scored per 90 minutes and AGVp90 for the most productive Bundesliga, Ligue 1, Premier League, LaLiga and Serie A players in the 2017/2018 and 2018/2019 seasons.

We will now discuss each of these in more detail.

IN-GAME CONTEXT

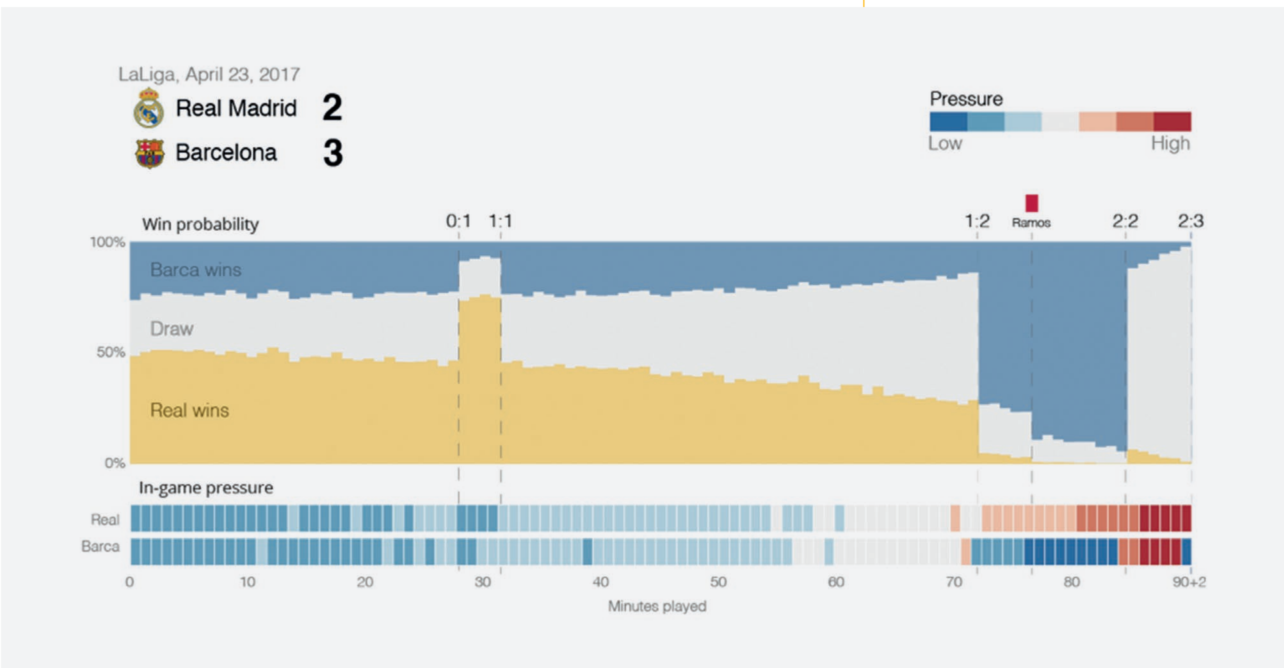
In the field of sports analytics, people have mainly focused on the effect of crucial game situations (often referred to as "clutch situations") with a presumed high level of mental pressure on performance. A clutch situation, according to Hibbs (2010), is 'a point in a competitive sport where the success or failure of the participants has a significant impact on the outcome of the contest.' Examples of such situations are Michael Jordan scoring with five seconds remaining to win the 1998 NBA Championship, Sergio Agüero's injury-time goal to win Manchester City's first Premier League title in over 40 years in 2012, and the fifth-set tiebreaker between Djokovic and Federer in the 2019 Wimbledon final.

While these are clear examples of clutch situations, questions remain over how to adequately define clutch performance, as well as the situations in which such performances occur. Typically, one defines clutch situations by hand (e.g. in the NBA, one defines the last five minutes of games separated by five points or fewer as clutch situations), and then computes performance metrics (both traditional and advanced) for players in such situations. These approaches are not directly transferable to football, however. Since up to 75% of all football games end with a goal difference below one, whether a team is truly under pressure or not often depends on a combination of multiple factors.

An alternative approach is to give a player a certain amount of positive or negative credit, according to how much their involvement advanced or reduced his team's chance of winning the game. Naturally, clutch situations will carry more weight, because those moments produce the largest swings in win probability. This approach was introduced as early as 1977 in baseball (Cramer, 1977), but only recently found its way into football, along with the first in-game win probability models. Such in-game win probability models estimate a sports team's likelihood of winning at any given point in a game, based on the current game state (i.e. the current score line, time remaining, red-carded players, relative team strength, and so on).

Based on an in-game win probability model, Robberechts et al. (2019) introduced the Added Goal Value (AGV) metric. This metric computes the average boost in expected league points due to a player's goals as the sum of the change in win probability multiplied by three and the change in draw probability. Figure 1 shows the relation between goals scored per 90 minutes and AGVp90 for the 2017/2018 and 2018/2019 seasons. The diagonal line denotes the average AGVp90 for a player with similar offensive productivity. Players below this line such as Neymar, Robert Lewandowski, Edinson Cavani and Luka Jovic have a relatively low added value per goal; while players above the line such as Robert Beric, Harry Kane, Paco Alcacer and Mauro Icardi add more value per goal than the average player.

This approach works well for goals and red cards, which have a measurable effect on win probability. Most other actions will have an infinitesimal effect. However, win probability can still be used to measure the 'crucialness' of game episodes if one looks at the difference in win probability between the current game state and two hypothetical game states where the home or away team has scored an additional goal. This approach was taken by Bransen et al. (2019) and is illustrated in Figure 2 for the game between Real Madrid and Barcelona on April 23, 2017. Barcelona's pressure level drops when they scored the 1-2 late in the game and Sergio Ramos was sent off. However, after Madrid's equaliser, the pressure level sharply increases again and only cools down after Messi's winning goal in the final minute of the game.



^
Figure 2.
Evolving pressure levels (bottom) and win probabilities (top) in Barcelona's 2-3 win against Real Madrid in the 2016/2017 La Liga 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.

PRE-GAME CONTEXT

To get an overall picture of the mental context, one cannot see the in-game situation apart from the context surrounding the match. Players will experience a different level of mental pressure ahead of a cup final, relegation battle or rivalry, compared to an end-of-season game with nothing at stake. This pre-game pressure will either fade or increase as the match progresses, depending on the in-game scenario.

The magnitude of the stakes and the importance of achieving success are the most important pre-game pressure facilitators. However, for the in-game context, there exists no clear definition of what constitutes a big game. Some people have solved this by focusing on a category of games that are undeniably "big", such as elimination games in NBA playoffs (Morgulev & Galily, 2018) or games in the latter stages of the UEFA Champions League (Dawson, n.d.). This approach, however, limits the number of games that can be analysed.

Bransen et al. (2019) used an alternative approach based on machine-learning to estimate the pre-game pressure of football games. While it is hard to provide a general definition for what makes one game more pressure-packed than another, football fans typically have good intuition about which of two games has the higher pre-game pressure. They initially asked football experts, therefore, to assess pairs of matches and state which one had higher stakes and subsequently used these rankings to train a machine-learning model. The model learns to reproduce the judgments of the experts by assigning scores to matches, so that a higher score is assigned to the match with the higher pre-game pressure. These scores define the pre-game pressure metric. To be able to rank games that were never rated by experts, the model describes each game using a set of general features that capture the following factors:

- **Team ambition:** Each team has different ambitions for the season – such as winning the league or simply staying up – that affect its pre-game pressure level. This ambition is captured 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 the 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 an estimation of how the current match's outcome will affect the probability that the team reaches a certain season outcome (e.g. avoiding relegation). This game importance is measured as the association between each possible game result (win-draw-lose) and each expected final league outcome (e.g. relegated, league champion) by simulating the remainder of the season.
- **Recent performance:** Football clubs are also subject to pressure based on recent form. Particularly for a big club, several consecutive poor performances will ratchet up the pressure. This is captured using the number of points obtained and the deviation from the expected performance over the previous 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, the match attendance, and how long ago the coach was appointed.

CONCLUSION

This section outlined the techniques that have been developed to value the actions that players perform. In particular, this section discussed how technical, tactical and mental aspects affect player performance and to what extent player performance metrics account for relevant contextual information. The existing metrics focus on evaluating player performance in a given context, but the important application of projecting player performance into a novel context (such as different tactics at a new club) has remained virtually unexplored to date and presents an interesting avenue for further research.


MESSI
10
22
unicef
Raku

REFERENCES

Aalbers, B., & Van Haaren, J. (2018). "Distinguishing Between Roles of Football Players in Play-by-play Match Event Data"; "Proceedings of the 5th Workshop on Machine Learning and Data Mining for Sports Analytics".

American Soccer Analysis (2020). "Goals Added: Introducing a New Way to Measure Soccer".https://www.americansocceranalysis.com/home/2020/4/22/37ucr0d5urxxtryn2cfhzormdziphq

Beal, R., Changder, N., Norman, T. D., & Ramchurn, S. D. (2020). "Learning the Value of Teamwork to Form Efficient Teams". *AAAI*, 7063–7070.

Bekkers, J., & Dabadghao, S. (2017). Flow Motifs in Soccer: What Can Passing Behavior Tell Us? *Proceedings of the MIT Sloan Sports Analytics Conference 2017*. MIT Sloan Sports Analytics Conference, Boston, USA.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., & Matthews, I. (2014). Win at Home and Draw Away: Automatic Formation Analysis Highlighting the Differences in Home and Away Team Behaviors. MIT Sloan Sports Analytics Conference 2014.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014a). Identifying Team Style in Soccer using Formations Learned from Spatiotemporal Tracking Data. 2014 International Conference on Data Mining Workshop.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014b). Large-Scale Analysis of Soccer Matches using Spatiotemporal Tracking Data. 2014 International Conference on Data Mining.

Bransen, L. (2017). Valuing Passes in Football Using Ball Event Data [Erasmus University Rotterdam]. https://thesis.eur.nl/pub/41346

Bransen, L., Robberechts, P., Van Haaren, J., & Davis, J. (2019). Choke or Shine? Quantifying Soccer Players' Abilities to Perform Under Mental Pressure. *Proceedings of the MIT Sloan Sports Analytics Conference 2019*.

Bransen, L., & Van Haaren, J. (2018). Measuring football players' on-the-ball contributions from passes during games. *Proceedings of the Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2018 Workshop*.

Bransen, L., & Van Haaren, J. (2020). Player Chemistry: Striving for a Perfectly Balanced Soccer Team. *Proceedings of the MIT Sloan Sports Analytics Conference 2020*.

Caley, M. (2015). *Premier League Projections and New Expected Goals*. https://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projection

Cowden, R. G., & Worthington, E. L. (2019). Overcoming Failure in Sport: A Self-Forgiveness Framework. *Journal of Human Sport and Exercise*; Vol 14, No 2DO - 10.14198/Jhse.2019.142.01. https://www.jhse.ua.es/article/view/2019-v14-n2-overcoming-failure-sport-self-forgiveness-framework

Cramer, R. D. (1977). Do Clutch Hitters Exist? *Baseball Research Journal*, 6, 74–79.

Dawson, A. (n.d.). We Analysed Who The Better Big-Game Player Is, Cristiano Ronaldo or Lionel Messi—And The Stats Show The Winner Is Clear. *Business Insider*. Retrieved August 6, 2020, from https://www.businessinsider.com/cristiano-ronaldo-or-lionel-messi-the-winner-is-clear-for-big-games-2018-4

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder Than Goals: Valuing Player Actions in Soccer. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Decroos, T., & Davis, J. (2019, September). Player Vectors: Characterising Soccer Players' Playstyle from Match Event Streams. ECML/PKDD, Würzburg, Germany.

Decroos, T., Van Haaren, J., & Davis, J. (2018). Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 223–232.

Decroos, T., Van Haaren, J., Dzyuba, V., & Davis, J. (2017). STARSS: A Spatio-Temporal Action Rating System For Soccer. ECML PKDD 2017.

Decroos, T., Van Roy, M., & Davis, J. (2020). SoccerMix: Representing Soccer Actions with Mixture Models. *Proceedings of the 2020 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2020, Ghent, Belgium.

Dick, U., & Brefeld, U. (2019). *Learning to Rate Player Positioning in Soccer*.

Fernández, J., Bornn, L., & Cervone, D. (2019). Decomposing The Immeasurable Sport: A Deep Learning Expected Possession Value Framework For Soccer.

Goes, F., Kempe, M., Meerhoff, R., & Lemmink, K. (2018). Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches.

Green, S. (2012). Assessing The Performance Of Premier League Goalscorers. https://opta.kota.co.uk/news-analysis/assessing-the-performance-of-premier-league-goalscorers/

Gregory, S. (2017). Re-Examining Finishing Skill. *Stats Perform*. https://www.statsperform.com/resource/re-examining-finishing-skill/

Gyarmati, L., & Stanojevic, R. (2016). QPass: A Merit-Based Evaluation of Soccer Passes. *ArXiv Preprint ArXiv:1608.03532*.

Hibbs, D. (2010). A Conceptual Analysis of Clutch Performances in Competitive Sports. *Journal of the Philosophy of Sport*, 37(1), 47–59. https://doi.org/10.1080/00948705.2010.9714765

Hill, D. M., Hanton, S., Matthews, N., & Fleming, S. (2010). Choking in Sport: A Review. *International Review of Sport and Exercise Psychology*, 3(1), 24–39. https://doi.org/10.1080/17509840903301199

Ijtsma, S. (2015). A Close Look at My New Expected Goals Model. 11tegen11. http://11tegen11.net/2015/08/14/a-close-look-at-my-new-expected-goals-model/

Knutson, T. (2018). *StatsBomb Data Launch—Beyond Naive xG. StatsBomb*. https://statsbomb.com/2018/05/statsbomb-data-launch-beyond-naive-xg/

Knutson, T. (2020). *StatsBomb Release Expected Goals with Shot Impact Height*. StatsBomb. https://statsbomb.com/2020/07/statsbomb-release-expected-goals-with-shot-impact-height/

Kothari, A. (2020). *XPo*. https://thecomeonman.github.io/xPo/

Kwiatkowski, M. (2017). Quantifying Finishing Skill. *StatsBomb*. https://statsbomb.com/2017/07/quantifying-finishing-skill/

Lawrence, T. (2015). Expected Saves. Deep XG. https://deepxg.com/2015/10/20/expected-saves/

Lawrence, T. (2018). Introducing xGChain and xGBuildup. https://statsbomb.com/2018/08/introducing-xgchain-and-xg-buildup/

Link, D., Lang, S., & Seidenschwarz, P. (2016). Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLoS ONE*, 11.

Liu, G., Luo, Y., Schulte, O., & Kharrat, T. (2020). Deep Soccer Analytics: Learning An Action-Value Function For Evaluating Soccer Players.

Llana, S., Madrero, P., & Fernández, J. (2020). The Right Place At The Right Time: Advanced Off-Ball Metrics For Exploiting An Opponent's Spatial Weaknesses In Soccer.

Lucey, P., Bialkowski, A., Mofort, M., Carr, P., & Matthews, I. (2015). "Quality Vs Quantity": Improved Shot Prediction In Soccer Using Strategic Features From Spatiotemporal Data. MIT Sloan Sports Analytics Conference 2015.

Mackay, N. (2017, July). Improving My 'xG added' Model – Mackay Analytics. Mackay Analytics. https://mackayanalytics.nl/2017/07/28/improving-my-xg-added-model/

Mackay, N. (2019). Introducing a Possession Value Framework. https://www.statsperform.com/resource/introducing-a-possession-value-framework/

Michalczyk, K. (2018, November 18). An Attempt To Extend Xg Gain. Kuba Michalczyk. https://kubamichalczyk.github.io//2018/11/18/An-attempt-to-extend-xG-gain.html

Morgulev, E., & Galily, Y. (2018). Choking or Delivering Under Pressure? The Case of Elimination Games in NBA Playoffs. *Frontiers in Psychology*, 9, 979. https://doi.org/10.3389/fpsyg.2018.00979

Perdomo Meza, D. (2017). Flow Network Motifs Applied to Soccer Passing Data. *Proceedings of MathSport International 2017*. MathSport International 2017.

Pollard, R., & Reep, C. (1997). Measuring the Effectiveness of Playing Strategies at Soccer. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(4), 541–550. JSTOR.

Power, P., Ruiz, H., Wei, X., & Lucey, P. (2017). "Not All Passes Are Created Equal": Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data. KDD 2017.

Robberechts, P. (2019). Valuing the Art of Pressing. Innovation in Football Conference 2019, London, United Kingdom.

Robberechts, P., Van Haaren, J., & Davis, J. (2019). Who Will Win It? An In-game Win Probability Model for Football. ArXiv:1906.05029 [Cs, Stat]. http://arxiv.org/abs/1906.05029

Rudd, S. (2011). A Framework for Tactical Analysis and Individual Offensive Production Assessment in Soccer Using Markov Chains. New England Symposium on Statistics in Sports.

Ruiz, H., Power, P., Wei, X., & Lucey, P. (2017). The Leicester City Fairytale?: Utilizing New Soccer Analytics Tools to Compare Performance in the 15/16 & 16/17 EPL Seasons. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1991–2000.

Russell, S., Jenkins, D., Rynne, S., Halson, S. L., & Kelly, V. (2019). What Is Mental Fatigue in Elite Sport? Perceptions From Athletes and Staff. *European Journal of Sport Science*, 19(10), 1367–1376. https://doi.org/10.1080/17461391.2019.1618397

Schweickle, M. J., Swann, C., Jackman, P. C., & Vella, S. A. (2020). Clutch Performance in Sport and Exercise: A Systematic Review. *International Review of Sport and Exercise Psychology*, 1–28. https://doi.org/10.1080/1750984X.2020.1771747

SciSports. (2020). Defensive Contribution Ratings in SciSports Platform Enable In-Depth Data-Driven Scouting of Defenders. https://www.scisports.com/defensive-contribution-ratings-in-scisports-platform-enable-in-depth-data-driven-scouting-of-defenders/

Singh, K. (2019, February). Introducing Expected Threat (xT) [Blog]. https://karun.in/blog/expected-threat.html

Trainor, C. (2014, October 21). Goalkeepers: How Repeatable Are Shot Saving Performances? StatsBomb. https://statsbomb.com/2014/10/goalkeepers-how-repeatable-are-shot-saving-performances/

Williams, A. M., Vickers, J., & Rodrigues, S. (2002). The Effects Of Anxiety On Visual Search, Movement Kinematics, And Performance In Table Tennis: A Test Of Eysenck And Calvo's Processing Efficiency Theory. *Journal of Sport & Exercise Psychology*, 24(4), 438–455. https://doi.org/10.1123/jsep.24.4.438

Yam, D. (2019, February 21). Attacking Contributions: Markov Models for Football. StatsBomb. https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/

Structure in football: putting formations into context

— Laurie Shaw[1]

1: Department of Statistics, Harvard University

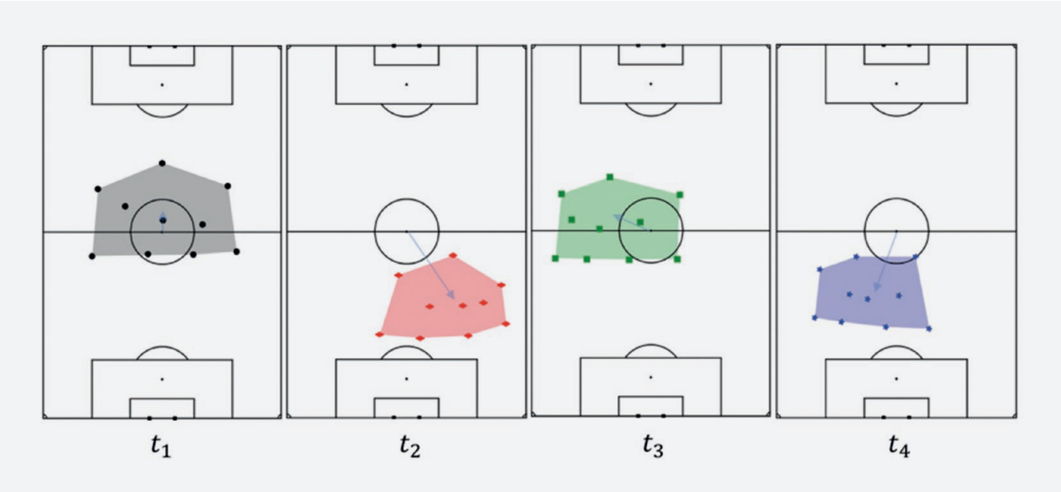
INTRODUCTION

Formations are the foundations of tactics in football. They provide structure to a team that helps the players to position themselves and broadly defines their specific roles in attack and defence. They are the means through which managers attempt to maintain control of high-value territory while denying their opponents the same (Fernandez and Bornn, 2018). The debate around formation tactics is as old as the game itself and is a central theme in the history of how the game is played. Innovations in formations have mirrored the evolving balance between defensive solidarity and attacking flair, discipline and freedom of expression and outcomes versus entertainment (Wilson, 2009). From the 2-3-5 'pyramid' in the late 1800s to the more balanced W-M formation of Herbert Chapman in the 1930s, the 4-2-4 of Brazil in the 1950s, the 5-3-2 of Helenio Herrera's Inter Milan in the 1960s, the zonal 4-4-2 of Arrigo Sacchi's AC Milan and the 4-2-3-1 of recent decades, formations have rich story of action and reaction.

In the modern game, formations are not rigid and unchanging; they are dynamic, adapting to the specific circumstances on the field, changes in personnel and each team's immediate objective. Managers frequently refer to the necessity of adopting different formations for different phases of the game and their importance to game management. Comprehensive analysis of an opponent must consider not only how a team is structured while defending their own goal, but also how they press higher up the field. In offense, it must consider their different configurations when playing the ball out from defence, progressing it up the field or attempting to break down a set defence in the final third.

The arrival of player tracking data has presented an opportunity to study team formations – and the transitions between formations – at an unprecedented level of detail. The data can reveal how the players have been instructed to organise themselves in different situations and against a variety of opponents. Detecting significant changes in formation during matches provides insights into how a coach reacts to certain game situations. With knowledge of the strategic framework within which players have been instructed to play we can attempt to understand their decision-making on an individual level, separating what have they been told to do and what do they do instinctively.

This chapter reviews some of the insights on formation tactics that data has revealed. In the next section we discuss the approaches that have been used to study formations in tracking data and highlight their findings. In the following section we present a case study that demonstrates how formations adapt to different game phases. In the final section, we review recent research into how teams might exploit knowledge of their opponent's formations to exploit weaknesses.



<
Figure 1.
The positions of the outfield players of the defending team at four different instants during a professional football match (shooting from bottom to top). The blue arrow indicates the average position of the team relative to the centre of the pitch.

CONSIDERATIONS AND SOLUTIONS

Measuring team formations with tracking data is a balancing act. At any particular instant a player may be a significant distance away from his or her formation position: covering for a teammate, chasing down an opponent, or making opportunistic runs into space. Formations must therefore be measured over a sufficiently long period of time that these deviations average out and we gain a more accurate understanding of each player's position.

However, player roles also vary with game phase: a full-back may be aligned with the centre-backs while defending in their own third of the pitch, but level with the forwards in the attacking third. The distinct phases of a game must be identified and formations measured in each phase separately. Furthermore, managers sometimes make wholesale tactical changes during a match: a complete change of formation (often accompanied by substitutions) to alter the flow of the match or close it out. We must detect these tactical changes and measure formations before and after them separately. So while tracking data should be aggregated over a period of time to average away the temporary departures of each player from their formation position, it must also be aggregated carefully, so that data from different game phases (or entirely different systems) are not mixed together, blurring the tactical picture.

A player's role within a formation is typically defined by their position relative to their teammates, rather than their absolute position on the field, particularly when defending. At any given instant, the area encompassed by the outfield players collectively is a relatively small fraction of the total area of the pitch: players move coherently as a group to maintain their spatial configuration. For example, Figure 1 indicates the positions of the defending team at four instants during the first half of a match. It is clear that, while the team occupies different areas of the pitch at each instant, the players largely retain their relative positioning, maintaining a 4-3-3 formation (four defenders, three central midfielders and three forwards).

Bialkowski et al. (2014) published one of the first quantitative analyses of team formations in football using tracking data (see also: Lucey et al. (2013), Bialkowski et al. (2016). They describe a role-identification methodology for measuring formations, iteratively refining estimates of the average spatial positions (and deviations from those positions) of 10 unique outfield roles throughout a match. At any given moment, player positions were measured relative to the average position of the team (as opposed to using their absolute positions on the pitch), accounting for co-ordinated team motions in the formation observations.

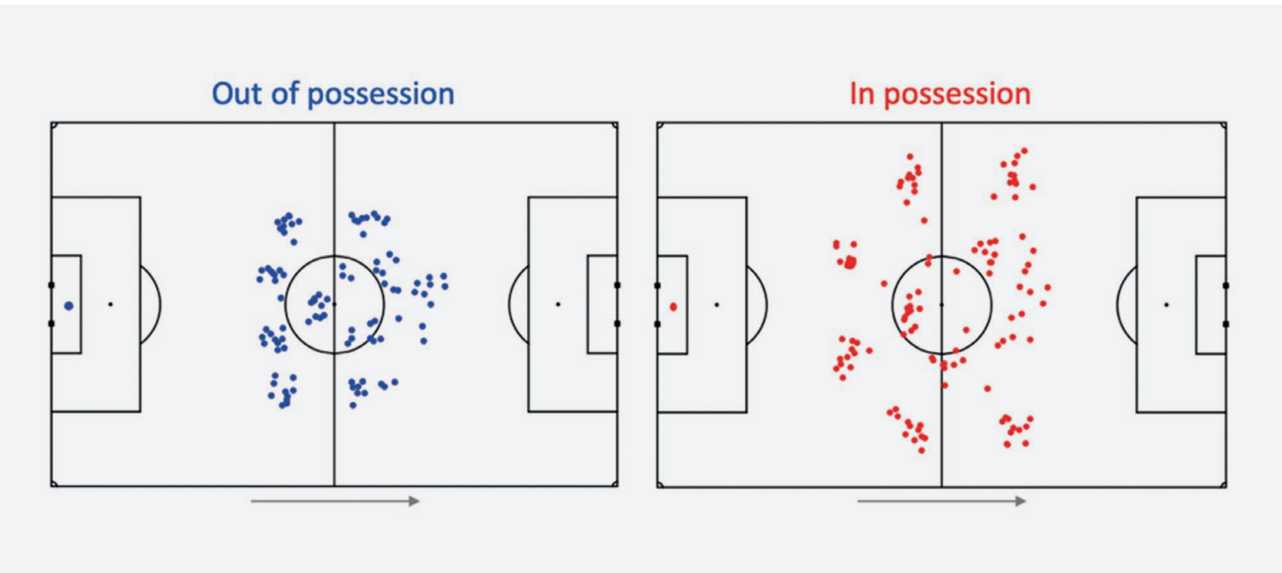
Bialkowski et al. (2014) applied their methodology to Prozone tracking data for a season of a 20-team professional league. A single formation observation was measured for each team in each half of every match; game phase information was not used and so the individual formation observations were a mixture of attacking and defensive configurations. Applying a clustering algorithm to their full set of formation observations, Bialkowski et al. (2014) identified 6 unique formation types: 4-4-2, 3-4-3, 4-4-1-1 and 4-1-4-1 are all visible in their results. To investigate how individual players exchanged roles throughout a game, they repeated their analysis, measuring formations in five-minute windows (rather than each half separately) throughout one of the matches in their sample. Their results showed how the midfielders – most notably the left and right wingers and the two central midfielders – exchanged positions throughout the match (see also Narizuka and Yamazaki, 2019).

In a follow-up paper, Bialkowski et al. (2016) extended their analysis to measuring formations in and out of possession separately for each team, finding in-possession formations to have a similar, although slightly more expansive, structure to those measured when teams were out-of-possession. However, when they measured team formations in five-minute periods and searched for distinct formation types within those observations, they found significant variations. Most notably, changes in formation coincided with the team's proximity to their opponent's goal, with the formations evolving to more aggressive configurations as the team advanced up the pitch. This was the first quantitative demonstration of the importance of game phase information to analysing team formations.

Shaw and Glickman (2019) presented a data-driven technique for measuring and classifying team formations as a function of game state, analysing the offensive and defensive configurations of each team separately and dynamically detecting major tactical changes during the course of a match. Using a season of tracking data from an elite professional league, they introduced a geometric approach to measuring formations, calculating the vectors between each pair of teammates at successive instants during a match and averaging these over a specified period of time to gain a clear measure of team formations in defence and attack. Defensive and offensive formations were measured separately by aggregating together consecutive periods of possession of the ball for each team into two-minute windows of in-play data, excluding periods of possession that lasted for less than five seconds (under the assumption that they are too short for either team to fully establish an offensive or defensive stance).

Figure 2 plots the full set of formation observations for one team during a single match (12 separate observations in possession and 9 observations out of possession). It is clear that, when out of possession (left plot) the team played with a 4-1-4-1 formation, with a single defensive central midfielder and a lone striker. In possession (right plot), the outside midfielders advanced to form a front three and the full backs moved level with the defensive midfielder. The right central midfielder played slightly deeper than the left central midfielder, introducing a small asymmetry to the team when attacking. While the relative positions of the defensive players in the team are clearly well constrained, the formation positions of the offensive players – particularly the striker – are much more broadly distributed, both in and out of possession, indicating greater freedom in their roles. Overall, the consistency of the observations indicates that the manager did not make a significant formation change during the match. Shaw and Glickman (2019) applied their methodology to identify a set of twenty unique formations that teams adopted over the course of a season. These unique formation types were used as templates to classify the formations used by teams during matches, study transitions between defensive and offensive configurations and detect major changes in formations during a match (see also Beernaerts et al. 2018, Müller-Budack et al. 2019).

Figure 3 provides examples of defensive and offensive formations that were frequently paired together by the teams in their dataset. The left-hand side of the diagram shows two defensive formations identified in the data, while the right-hand side shows three offensive formations. The links between them indicate the formations that were regularly combined as possession was gained and lost.



^
Figure 2.
The full set of formation observations for one team throughout an entire match. The left plot indicates 9 defensive formation observations, the right plot indicates 12 offensive formation observations.

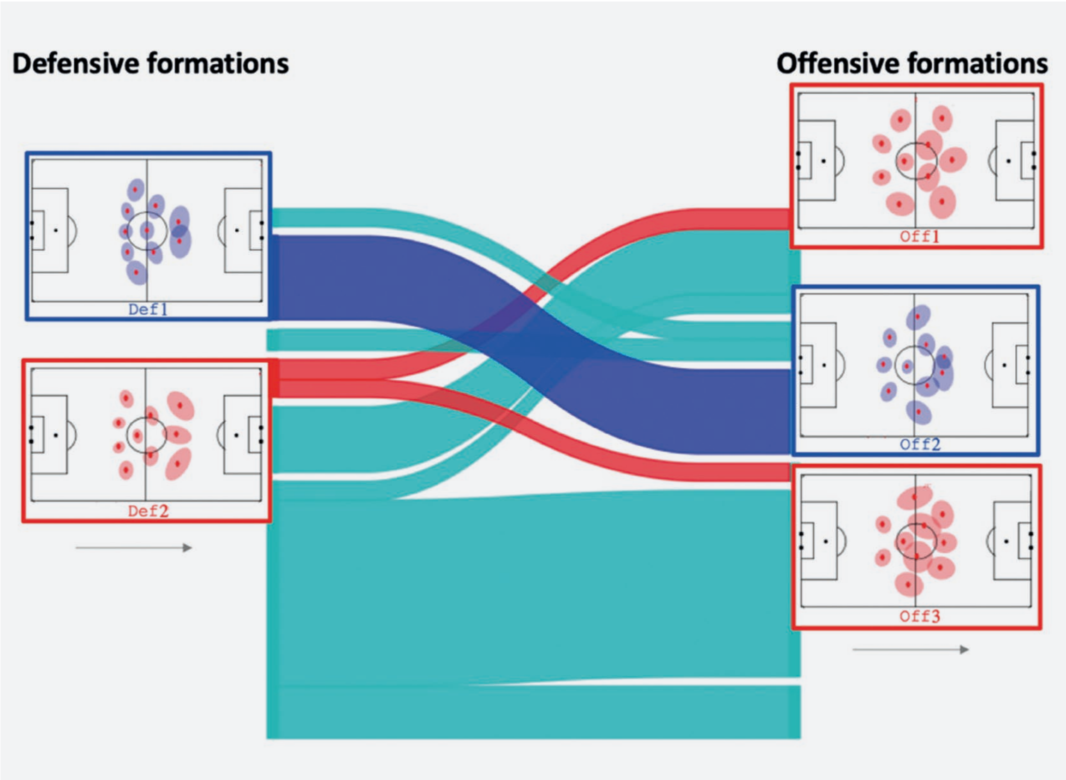


Figure 3.
Two examples of the typical pairings between defensive and offensive formations in a sample of 180 matches. All formations are orientated to shoot from left to right.

The example highlighted in blue indicates that teams that defended using the formation *Def1* typically transitioned to the formation *Off2* when in possession of the ball. The relationship between the two formations is clear: the outside defenders, or wingbacks, advanced when the teams gained possession and the two outside midfielders tucked in behind the two forwards.

The second example, highlighted in red, demonstrates that teams defending using *Def2* would transition into either *Off1* or *Off3* when they gained possession. In the former, the outside forwards pushed wide and the full-backs advanced into midfield, whereas in the latter the front three remained narrow with the full-backs advancing further up the field to provide width to the team. These examples show that some defensive configurations seem to give more flexibility in terms of attacking options than others.

PHASES	PHASE TYPES
Offence	Ball retention Progression Chance creation
Defence	Low block Mid block High block
Transitions	Establishing possession Counterattack Counterpressing Fall back
Set Pieces	Corners, free kicks, goal kicks & throw-ins

<
Table 1. Example of game phase categorisation (based on discussions with analysts at the German Football Federation).

GAME PHASES

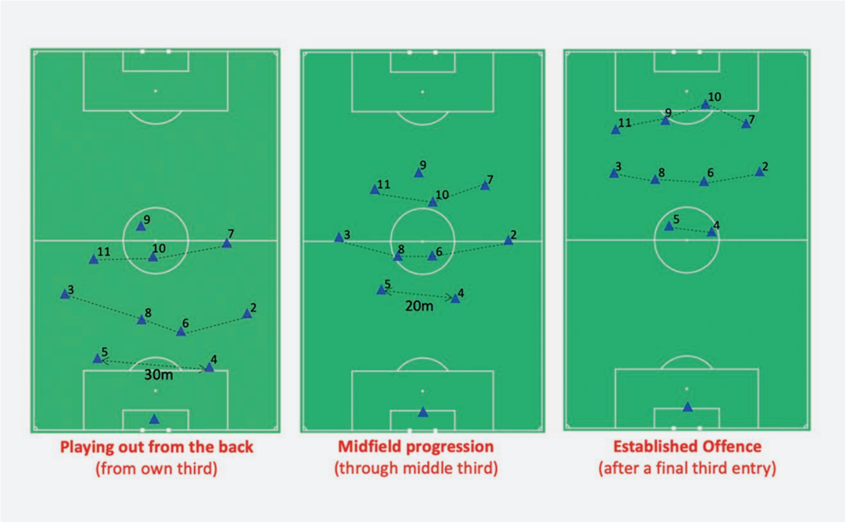
The results discussed above reinforce the notion that a team's formation will vary based on the *game phase*. The concept of game phases derives from the idea that any moment of a match can be categorised based on the immediate intentions of each team. While there is no universally accepted definition of games phases, an example is shown in Table 1.

In this example, matches are broken down into four phases – Offence, Defence, Transitions and Set Pieces – each of which consist of a number of phase types. The Offensive phases categorise periods of possession into ball retention (ensuring possession of the ball is maintained), progressing the ball towards the opponent's goal, and active chance creation. Defensive phases effectively indicate how far up the pitch the team is attempting to defend. The Transition phases deal with the intentions of a team in the moments that follow an exchange of possession. Set pieces label all dead ball situations (and could be further refined into 'first ball', 'second ball', and so on).

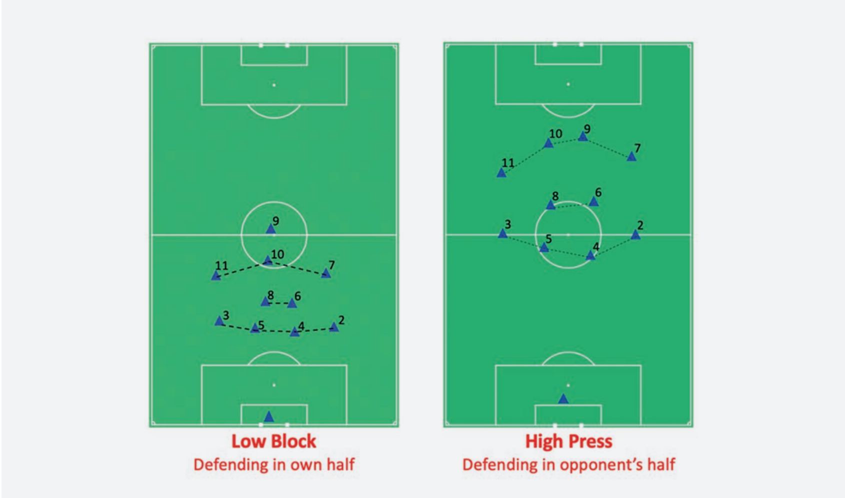
Automated detection of game phase is a challenging technical problem and many current categorisations rely on human analysts' interpretation rather than an algorithm. As game phases are loosely related to where a team is located on the pitch, a simple method is to base the definitions of game phases on the distance of the team from their own goal. In Figures 4 & 5, we show formations measured in different game phases for a major team in a cup semi-final.

Figure 4 shows formations measured in three simple phases of possession based on the vertical position of the team's centroid: defensive third, middle third, and final third. While playing the ball out from their own third of the pitch, the two central defenders (#4 and #5) were typically separated by a distance of 30 meters, with the midfielders (#6 and #8) dropping deep and the full backs pushed further up the field. The structure is similar in midfield progression (middle plot), with the gap between the two central defenders narrowing to 20 meters. In the final third, the formation changes significantly, resembling a 2-4-4: the trio of attacking midfielders (#11, #10 and #7) advance level with single striker (#9) to form a single attacking line, with the central midfielders and full-backs forming a second line behind them.

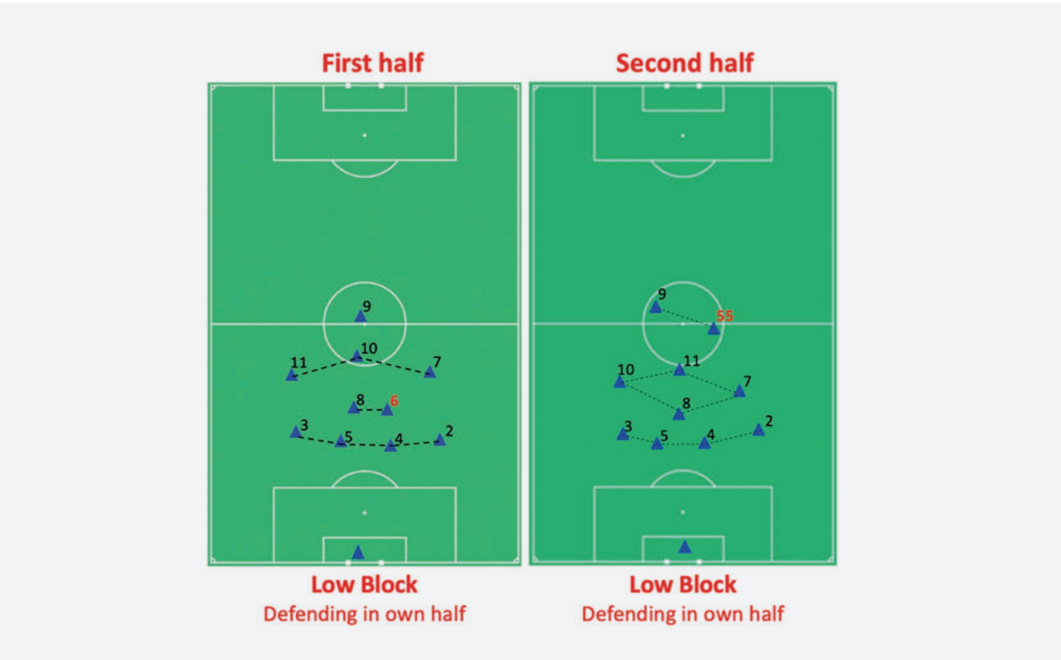
Figure 5 shows the formations in two defensive phases: low-block (defined as periods when the team were defending in their own half) and high-press (defined as periods when the opponents were attempting to play the ball out of their defensive third). The low block is clearly a compact 4-2-3-1, with the trio of attacking midfielders remaining in advance of the two central midfielders and the back four positioned about ten meters outside their own D. In the high press, the data indicates that their formation resembled a 4-2-4 (or an aggressive 4-4-2), with the front four forming a single line and the two central midfielders sitting in front of the back four. Their opponents played with a 4-3-3 while in possession in their own third, with a single midfielder playing deep and two midfielders more advanced on either side of him. Players 6 and 8 may therefore have remained deep to help protect the defence



^
Figure 4.
Team formations in 3 basic phases of possession (first half only).



^
Figure 5.
Team formations in 2 basic phases of defence (first half only).



<
Figure 6
The impact of the half-time change in formation on the team's low block.

against the front three should their opponents break through the first line.

Across the five game phases depicted in Figures 4 and 5, the players were configured in four related but distinct formations: 2-4-3-1, 2-4-4, 4-2-3-1 and 4-2-4. The clear structure to each measurement emphasises how formations form the backbone of team tactics, while the changes from one phase to another demonstrate how the team adapted to different situations.

The formations shown in Figures 4 and 5 were generated using tracking data from only the first half of the match. Once the entire game has been broken down into the constituent phases, the tracking data for each phase can be aggregated to create a single formation observation for each phase. However, before the data is aggregated, it is necessary to check whether there was a major formation change at some stage of the game.

In this case study, the team made a significant tactical change at half time. Figure 6 demonstrates how their formation in the low block changed from the 4-2-3-1 to a midfield diamond. Players 10 and 11 exchanged positions and, after 15 minutes of the second half, player 6 was substituted for player 55.

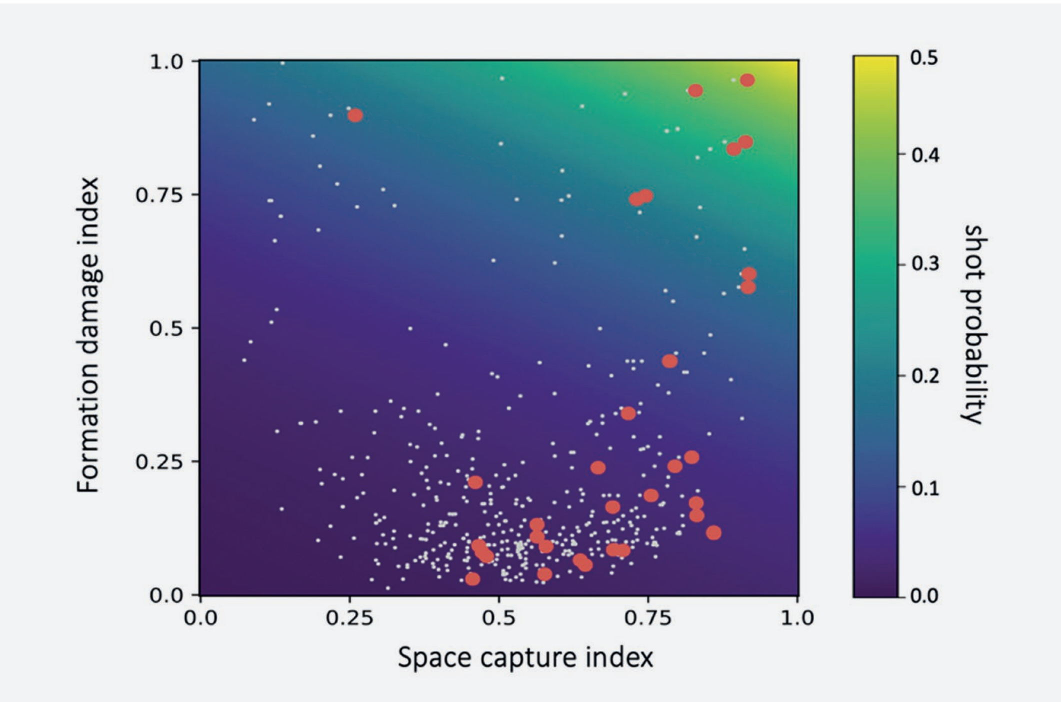
How did this formation change affect the flow of the game? Their opponents had been the superior team in the first half, but the second half was more balanced, both in terms of possession and chance creation. It is worth noting that, over the course of the season, the coach frequently changed his team's formation during matches. Formation detection algorithms can be used to flag interesting tactical changes made by a manager over several seasons that can then be studied to anticipate how he or she may react in the future.

FORMATION DISRUPTION

One of the immediate benefits of measuring formations is that the concept of tactical discipline can then be quantified. This is particularly pertinent to defensive organisation: when does the defensive shape of an opponent become disrupted? Football is a territorial game and formations are a strategic tool for ensuring that high-value territory is well-guarded. To create goalscoring opportunities, the attacking team must attempt to occupy space near their opponent's goal long enough to produce a clear shot at goal. One way to achieve this is to create disorder in the defensive system by manipulating defenders away from the spaces they defend.

Sormaz & Nichol (2019) explored the relationship between formation disruption and space creation, establishing a link between off-the-ball runs, defensive disorganisation and shooting opportunities (see also Memmert et al. 2017). They introduced a new metric, formation damage, that quantifies the degree to which the positions of the defending players have deviated from a reference formation (which was inferred from the data). The space created by off-the-ball runs – defined as periods of sustained acceleration in the opponent's half – was quantified by measuring the maximum area controlled by a player over the duration of their run using Voronoi Tessellation.

Figure 7 shows that runs that both capture space and help to disrupt an opponent's formation are more likely to lead to chance creation. The probability that an attacking play will produce a shot – as indicated by the colour scale – increases towards the upper-right hand corner of the plot, corresponding to off-the-ball runs that simultaneously capture large areas in the opponent's half and damage their formation by forcing defenders out of position. The method introduced by Sormaz and Nichol (2019) provides a metric for assessing run quality: runs that achieve high scores in both metrics are related to an increase

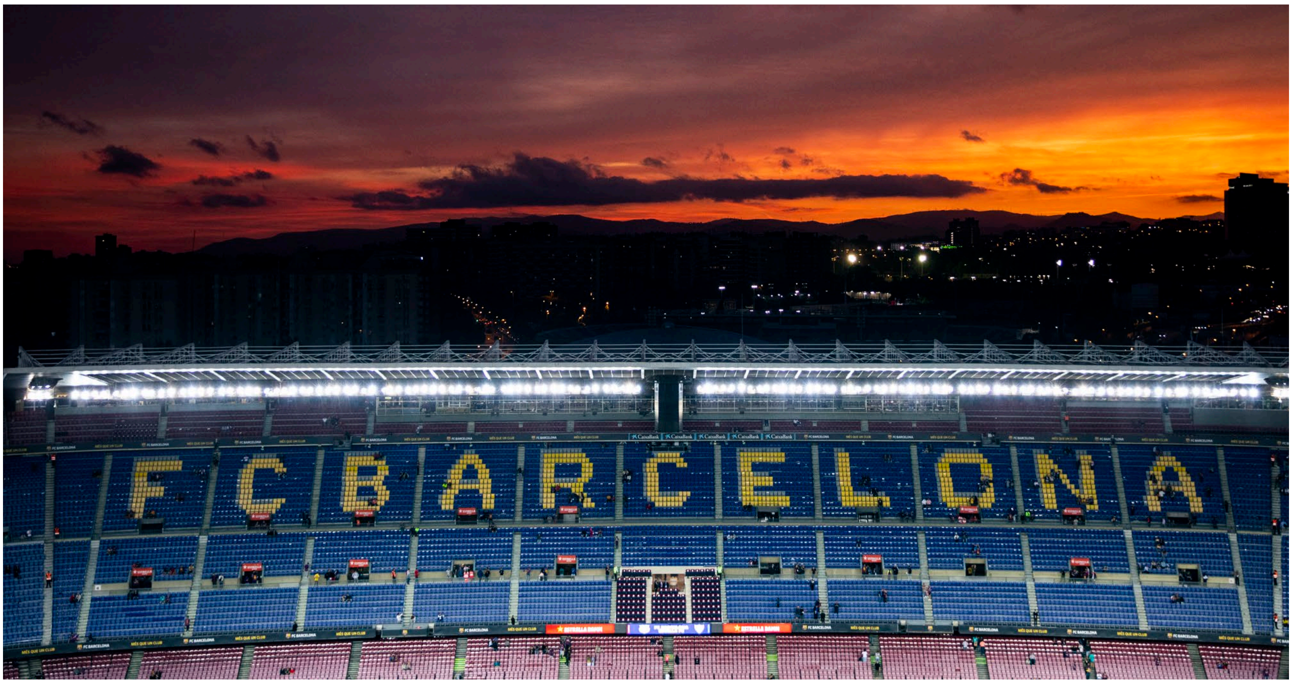


<
Figure 7
Space captured by off-the-ball runs made by attacking players in the opposing half (x-axis), plotted against the degree of formation damage that occurred during these runs (y-axis). Small white dots represent individual runs made players in the data; larger red dots highlight runs made in an attacking play that produced a shot on goal. The colour scale indicates the results of a logistic regression of shot probability on space creation and formation damage (brighter colours indicate a higher shot probability). Taken from Sormaz and Nichol, 2019, with permission from the authors.

in attacking threat. A natural next step would be to demonstrate a clear causal link between player movement and formation damage – identifying the types of runs that drag players out of position – and to consider the disruptive effects of counterpressing.

Llana et al. (2020) demonstrate how data can be used to identify strategic weaknesses in the positioning of defending players. Using the methodology outlined in Shaw & Glickman (2019), they allocate zones to individual players in the defending team: each player is assumed to be responsible for covering his or her own unique zone. They then find instances in which the attacking team passed the ball into one of these zones whilst bypassing the defending player responsible for guarding it (who is assumed to be 'out of position'). Llana et al. (2020) demonstrate how these situations cause other players in the defending team to move out of position in turn, propagating formation disruption throughout the team and creating space for the opponents. Video examples of these situations can be found at https://bit.ly/2PIBeV2.

Using the Expected Possession Value framework presented by Fernandez et al. (2019), Llana et al. (2020) quantified the cost of a given player being caught out of position by calculating the probability that a pass into their zone would increase the chances of the possession resulting in a goal. In a case study of the UEFA Champion's League group stage match between Tottenham Hotspur and FC Barcelona in the 2018-19 season, they showed how their methods revealed the high-value nature of Barcelona passes played beyond Tottenham's full-backs Trippier and Davies and into the zones they were expected to cover.



CONCLUSION

The discussion of formations is as old as the game itself, but data-driven formation analysis is still a relatively young field. With the increasing availability of tracking data, new methods for measuring formations in various match contexts are now being developed. Studies of formation discipline and the disruption of defensive blocks are providing clear demonstrations of how these methods can be used for opposition analysis and match preparation.

How might analytics influence the thinking of the next generation of coaches? Analysis of tracking data has demonstrated that there is much greater variety in the formations utilised by teams – even within a single game – than is discussed in standard pre- and post-match reports. In future, data will help to reveal answers to important questions: what are the typical strengths and weaknesses of different formations (especially when pitted against one another)? How do formations affect playing style, and do they enhance or supress a particular player's abilities? Formations also influence the dynamic elements of a player's role – the opponent's they mark, the space they create or defend, the co-ordinated runs made by players; data can also be used to explore these aspects of team strategy.

For those of us who love to dissect and study the beautiful game, the different perspectives that data analysis can provide will hopefully lead to new insights that, when combined with the knowledge and experience of leading coaches, will help to drive the next steps of tactical innovation in football.

REFERENCES

Wilson, J. (2009). *Inverting the Pyramid.* Nation Books.

Beernaerts, J., De Baets, B., Lenoir, M., De Mey, K., Van de Weghe, N., (2018). Analysing Team Formations in Football With the Static Qualitative Trajectory Calculus. *Proceedings of the 6th International Congress on Sport Sciences Research and Technology Support - Vol 1*: icSPORTS, 15-22

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., and Matthews, I., (2014) Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. 2014 IEEE International Conference on Paper Presented at the Data Mining (ICDM). 14–17 Dec 2014

Bialkowski, A., Lucey, P., Carr, P., Matthews, I., Sridharan, S., and Fookes, C. (2016) Discovering Team Structures in Soccer from Spatiotemporal Data. *IEEE Transactions on Knowledge and Data Engineering*, 28(10), pp. 2596- 2605.

Fernandez, J. and Bornn, L., (2018), Wide Open Spaces: A Statistical Technique for Measuring Space Creation In Professional Soccer. Sloan Sports Analytics Conference, retrieved from http://www.sloansportsconference.com/wp-content/uuploads/2018/03/1003.pdf

Fernandez, J., Bornn, L., and Cervone, D.,(2019), Decomposing the Immeasurable Sport: A Deep Learning Expected Possession Value Framework For Soccer. Sloan Sports Analytics Conference, retrieved from http://www.sloansportsconference.com/wp-content/uploads/2019/02/Decomposing-the-Immeasurable-Sport.pdf

Llana, S., Madrero, P., and Fernández, J., (2020). The Right Place at the Right Time: Advanced Off-Ball Metrics for Exploiting an Opponent's Spatial Weaknesses in Soccer. Sloan Sports Analytic Conference, retrieved from http://www.sloansportsconference.com/wp-content/uploads/2020/02/Llana_The-right-place-at-the-right-time.pdf

Lucey, P., Bialkowski, A., Carr, P., Morgan, S., Matthews, I., and Sheikh, Y., (2013). Representing and Discovering Adversarial Team Behaviors using Player Roles. CVPR, 2013.

Memmert, D., Lemmink, K. A. P. M., Sampaio., J., (2017). Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine*, 47, 1. DOI 10.1007/s40279-016-0562-5

Müller-Budack, E., Theiner, J., Rein, R., & Ewerth, R. (2019). "Does 4-4-2 Exist?" An Analytics Approach to Understand and Classify Football Team Formations in Single Match Situations. *Proceedings of the 2nd International Workshop on Multimedia Content Analysis in Sports* (pp. 25-33).

Narizuka, T., Yamazaki, Y. (2019). Clustering Algorithm for Formations In Football Games. *Sci Rep 9*, 13172 (2019). https://doi.org/10.1038/s41598-019-48623-1

Shaw, L., and Glickman, M., (2019). Dynamic Analysis of Team Strategy in Professional Football. *FC Barcelona Innovation Hub Sports Analytics Summit*, pp. 1–13. Retrieved from: https://static.capabiliaserver.com/frontend/clients/barca/wp_prod/wp-content/uploads/2020/01/56ce723e-barca-conference-paper-laurie-shaw.pdf

Sormaz, M., and Nichol, D., (2019). Quantifying the Impact of Off-The-Ball Movement in Football. OptaPro Analytics Forum 2019. Talk retrieved from https://www.youtube.com/watch?v=IG6LJo-5c_6U.

6 Defensive organisation disruption for team success

— Matthias Kempe[1], Floris Goes[1], Mat Herold[2] & Pascal Bauer[3]

1: Department of Human Movement Sciences, University of Groningen, University Medical Center Groningen (UMCG), Groningen, The Netherlands

2: Institute of Sports and Preventive Medicine, Saarland University, Saarbrücken, Germany

3: German Football Federation (DFB)

INTRODUCTION

Football can be considered a simple game between two teams of 11 players, each with the task of putting the ball in the opposing team's goal. As ball possession changes constantly, the game can often look rather chaotic. However, as students of the game, we know better. Within this chaos, patterns derive from creative interactions of players or sequences of actions learned during hours of training sessions instructed by the coaching staff. These patterns are described in the community as tactics or tactical behaviour. Tactical expertise and the ability to teach tactical behaviour is, rightfully, seen as the key expertise of coaches (Brink, Kuyvenhoven, & Toering, 2018) with some of them labelled as 'tactical wizards'. The task of the supporting staff is to help these wizards fine-tune their tactical approach and select the best fitting players. To do so, it is important to understand and quantify tactical behaviour. Therefore, the interaction and organisation of the players within a team and subunits of one team should be considered in relation to the opponent. (Gréhaigne, Bouthier, & David, 1997)

Within this section, we would like to introduce a new measure to further understand and quantify the tactical behaviour of teams and players called defensive disruption. Furthermore, we explain the origins and theoretical underpinning of this idea. In doing, so our goal is to highlight that the understanding of complex issues, like tactical behaviour, is best explored via a team effort of practitioners and sport (behaviour) and data scientists.

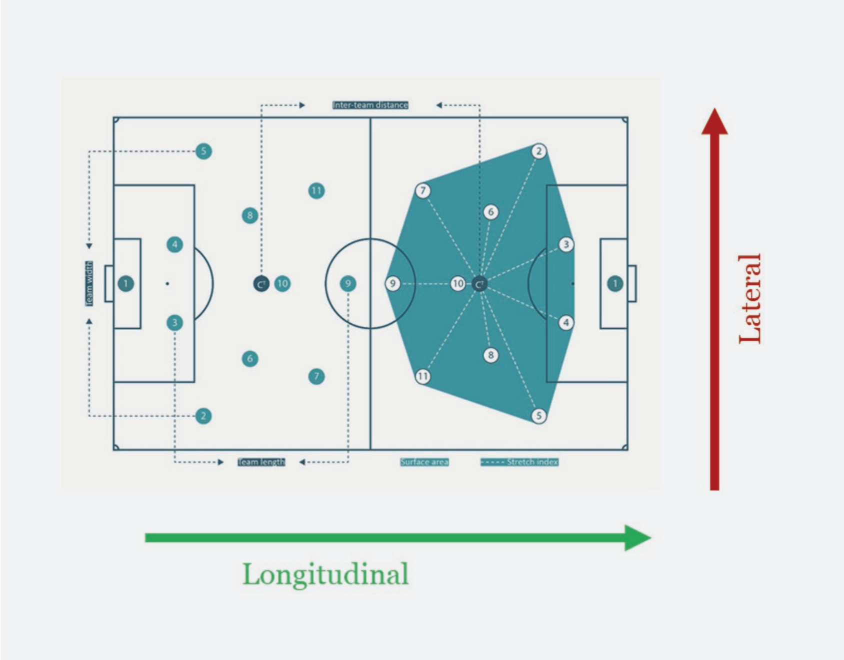
UNLOCKING THE POTENTIAL OF DATA SCIENCE TO SUPPORT TACTICAL PERFORMANCE ANALYSIS

Within two review papers that summarise approaches used by sport and computer science to study tactical behaviour in recent decades, we observed several conceptual differences (Goes et al. 2020; Herold et al. 2019). Both domains have contributed distinctly different studies, with the first being more focused on developing theories and practical implications, and the latter more on developing techniques.

Computer or data science concerns itself with the theoretical foundations of (computationally retrieving) information, typically yielding advanced analyses and high-level representations of large and complex data (Gudmundsson & Horton, 2017). Although not aimed at practice, these studies still hold valuable contributions, as explorative techniques such as subgroup discovery (Grosskreutz & Rüping, 2009) have the benefit that patterns can be discovered based on how 'interesting' they are without consciously searching for them. This enables data-science approaches to find and calculate spatial and temporal aggregates that would not have been thought of by sport scientists. We visualised some of the commonly used are spatial aggregates like team centroid, stretch index, or team surface in Figure 1.

In general, we see an added benefit of a synergy between both domains. The studies by Link et al. (2016), Rein et al. (2017), and Goes et al. (2019), are examples of sports science work that utilises observational designs in which large datasets were used for the development and validation of new features that assess tactical performance (Kempe, Meerhoff, & Lemmink, 2018; Link, Lang, & Seidenschwarz, 2016; Rein, Raabe, & Memmert, 2017).

On the other side of the spectrum, studies by Power et al. (2017), Spearman



^
Figure 1
Commonly used spatial aggregates to describe team tactical behaviour; CT- Team Centroid, Surface Area as the highlighted blue area.

et al. (2017), Andrienko et al. (2017), and Fernandez and Bornn (2018) can all be regarded as examples of studies that predominantly involved expertise from computer and data science, but also involved domain expertise from sports (science). To better combine the efforts of sport science, data science and practice, we provide a framework for optimising this collaboration (Figure 2).

Within our framework, we propose that collaboration between sports science and computer science should be cyclical rather than parallel. Both domains should work in close contact, preferably in teams, combining their approaches instead of working in a segregated fashion. By applying techniques from computer science to sports science research designs, sports scientists could arrive at new perspectives on performance and more in-depth answers to their research questions. The other way around, research questions deduced from theory and observation by sports science can be used by computer science to define the scope of their search and the development of appropriate technologies to derive information from position tracking data.

The combination of sport science and computer science expertise can also help to solve the problem of transferring scientific results into practice. As tactical behaviour is highly dependent on the context (Gréhaigne, Bouthier, 1999; Rein & Memmert, 2016), larger real-life datasets collected in actual competitive matches in combination with methodology that enables capturing complex patterns might allow one to

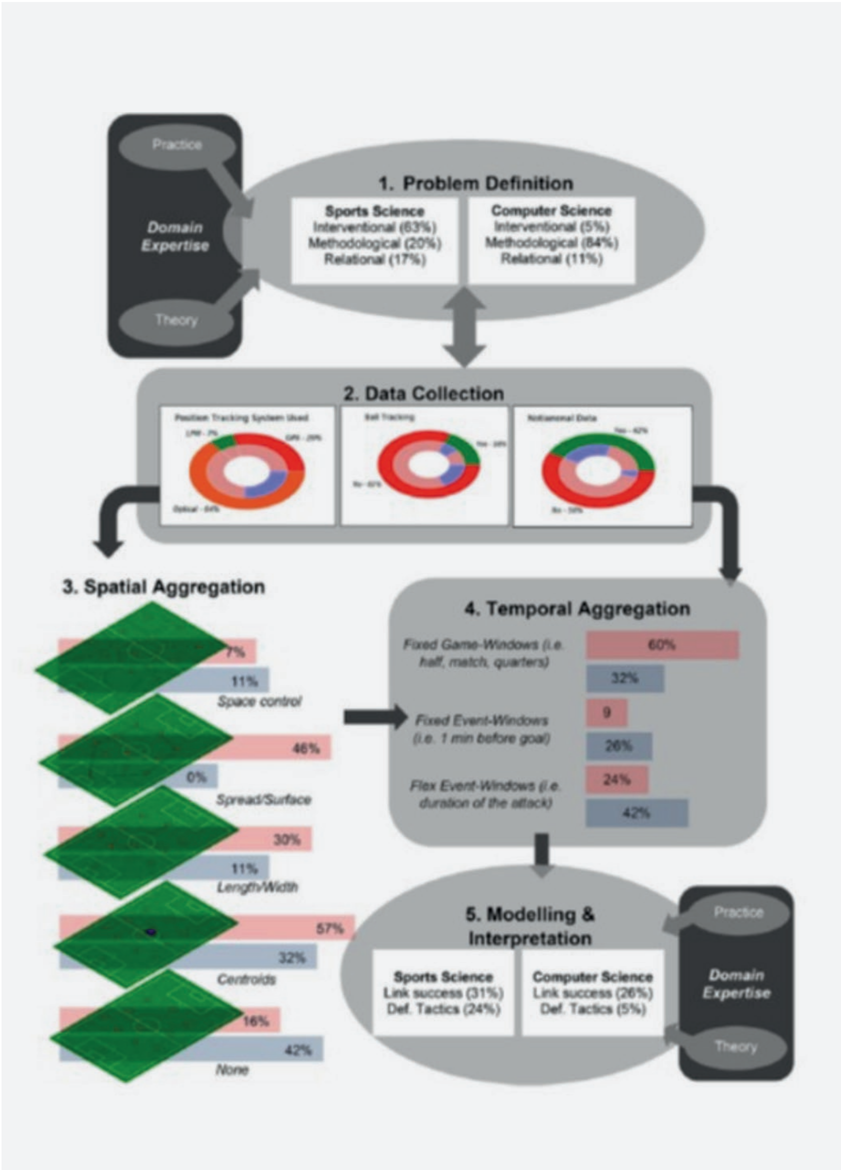
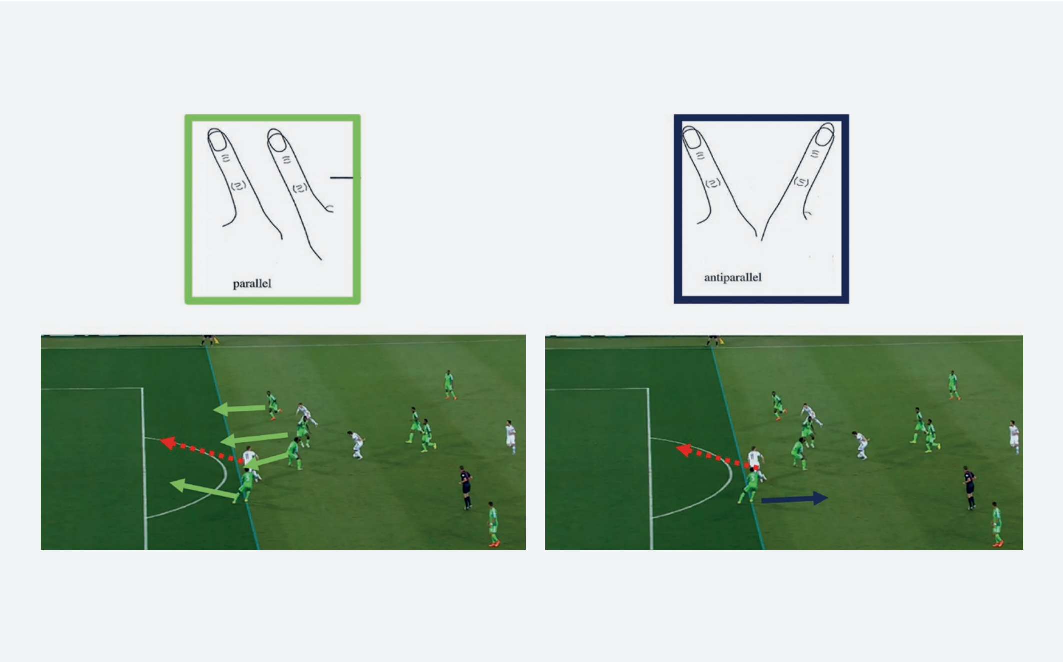


Figure 2. Conceptual framework for the combination of sports science (translucent red bars) and computer science (translucent blue bars) expertise in the study of tactical behaviour in football. Based on the results from the current systematic review. Bars with percentage represent the relative occurrence of a certain method or feature within a domain. Abbreviations: SSG, Small-Sided Games; LPM, Local Position Measurement. The panel was adapted from work of Goes et al. (2020)

draw conclusions about performance with a stronger ecological validity.

Ultimately, spatial features – no matter their complexity – hold little meaning when aggregated over a full match, and temporal aggregation is essential to place spatial behaviour in a temporal context. Most sports science studies are aggregated over fixed windows independent of game-context, like a match or half, which limits interpretability. We argue that deriving meaning from spatial features requires the use of event-based time-windows. This is more common in computer science studies, as using event-based time-windows allows one to draw conclusions about for example a pass, dribble, or set-piece. On such a small timescale, it is much easier to find structural patterns than on the level of the entire game.

In summary, sport science, data science, and practice should work in teams with a cyclical collaboration in order to solve relevant complex issues using large datasets broken down using spatial features on an intelligent temporal scale.



<
Figure 3
Exemplary behaviour of the defensive line to react through pass of the attacking team (red line) in alignment with the finger wagging experiment of Kelso, by a) moving in parallel (green lines) or b) anti-parallel (blue line).

TACTICAL BEHAVIOUR AS A COMPLEX DYNAMIC SYSTEM

In the next section, we'd like to introduce the theoretical underpinning of our defensive disruption measure. It is based on the idea that in order to understand tactical behaviour, one needs to understand the organisation of the players in space and time. Players constantly need re-organise themselves and adapt their positioning and movements to stay in the planned organisational structure. Most researchers agree this complex organisation could best be understood using the complex (dynamic) systems approach (Balague, Torrents, Hristovski, Davids, & Araújo, 2013). This approach provides a framework of ideas, equations (non-linear), and variables to understand the coordination processes in football (Davids, Araújo, & Shuttleworth, 2005). One of the main assumptions of this approach is that although its parts (players) are constantly moving in a rather chaotic way, the system (team) itself tends towards a stable (self-organised) structure. In reaction to constraints (i.e. score-line, attacking vs. defending) outside of the systems, the various parts (self-) organise themselves to create an optimal stable pattern accounting for these constraints as fast as possible.

As highlighted by the research by Kelso et al. (Haken, Kelso, & Bunz, 1985) with their famous finger-wagging experiment, these stable movement patterns are either parallel (in phase), meaning all parts (players) moving in the same direction at the same time, or antiparallel (anti-phase), moving in the opposite direction at the same time. A simplified example of this idea is given in Figure 3, using the defensive line and the attackers of the opposing team as a "stable" system. As visualised in this figure, the defenders can organise their movements to fulfil their shared task (preventing a goal) given the constraints (offside rule, movement of the attackers, the position of the attackers) at the time.

THE GOAL IS DESTABILISATION

Using this theoretical background leads to two hypotheses science and practice agree on:

- The defending team wants to stay in a stable organisation and minimise the time of transition between unstable and stable phases.
- Hence, attacking teams will try to disrupt the defensive organisation and try to create long passages of unstable defensive organisation.

Several studies in the last 20 years have shown that, in principle, football is an in-phase sport, in which teams tend to move up and down the field as well as side to side in synchrony and in the same direction (Duarte et al. 2012; Siegle & Lames, 2013). Analysis of variables like the mean team position (team centroid) has demonstrated that inter-team interaction is strongly synchronised, especially in the longitudinal direction (Frencken, Lemmink, Delleman, & Visscher, 2011). However, research with team level variables like the team centroid failed to find evidence of an inter-team synchrony disruption in moments before key events like goals or shots. One might argue that team level variables are not specific enough to study the tactical behaviour that characterises successful and non-successful attacks, and the use of subgroup-level variables is therefore recommended (Bartlett, Button, Robins, Dutt-Mazumder, & Kennedy, 2012). Following this recommendation, we studied the behaviour of subgroups (defensive, midfield, and attacking line) within one team (intra-team) and between opponents (inter-team) in 118 Dutch Eredivisie matches (Goes et al. in revision).

LARGE-SCALE SPATIOTEMPORAL ANALYSIS OF DYNAMIC SUBGROUPS USING POSITION TRACKING DATA

Within this study, we found that opposing teams (see Figure 4, Centroids in longitudinal (Cx) and in latitudinal (Cy) direction) as well as opposing subgroups mostly act in-sync with timely close coupling. This means that for the most part, attacking and defending teams don't have an offset in their movement. The only subunits that showed varying behaviour are the defensive line of the defending team and the attacking line of the attacking team. This asynchronous behaviour was characterised by the fact that when the attackers moved towards the opposing goal, the defenders did not fall back to their own goal but rather moved towards the attacking players. During this behaviour, strong timely coupling was also apparent.

As explained in the example in Figure 3, if destabilising the defence would be a measure for success in football, we should find a decoupling (time offset in the movement) of the attacking and defending teams, or at least in their subunits, in successful possessions. To operationalise success, we computed the potential for every reception to result in a scoring opportunity, using a zone value (Z) for every ball reception during an attack. To do so we used Link's 'dangerousity' (Link et al. 2016), in which every reception is awarded a score between 0 – 1 (0- no change of goal scoring, 1- goal). Based on the location of the reception in relation to the goal we deduced points for the defensive pressure on the ball receiver. This concept has been validated in work by Goes, Kempe, and Lemmink (2019, 2020).

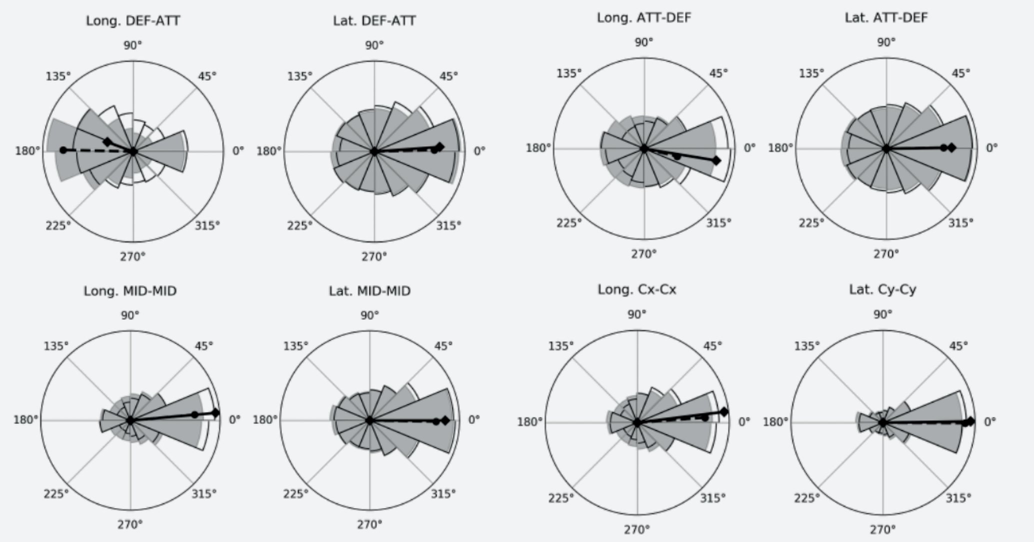


Figure 4. Rose plots of relative phase distributions for inter-team variables. Data is grouped in 22.5^0 bins, in which the radius of the bin represents the relative occurrence. Grey bins with no edges represent non-successful attacks, while white bins with black edges represent successful attacks. Black dotted lines with circular markers represent the mean direction **θ** and mean resultant length R of the non-successful distributions, while black solid lines with diamond markers represent those of the successful distributions. The panel was adapted from work of Goes et al. (in revision)

If a single possession reached a value above the threshold (>0) it was classified as successful as it allowed for the direct creation of a scoring opportunity.

When comparing successful (N = 1.237) and unsuccessful (N= 11.187) possessions we found a decoupling in the global team behaviour as well as in their subunits. However, these effects were mostly marginal and could be explained by a large sample size. On the other hand, the decoupling of the defending team's defensive line and the attacking line of the attacking team indeed yielded a pronounced effect with a mean offset of 19^0 (see Figure 4 the differences between the dotted and the straight line in the highlighted rose plot).

These results finally provide empirical evidence to the community that decoupling and therefore, destabilising the defensive organisation of the opposing team leads to success in football.

As described earlier, the basic assumption of our recent paper (Goes et al. in revision) was already established 20 years ago. However, only now with the help of player tracking data and data science approaches to automatic detect team subunits have we been able to prove it.

NOT EVERY PASS CAN BE AN ASSIST

The previous paragraphs and overview articles provided us with a framework of how we like to conduct our research on the evaluation of tactical performance. We translated these ideas into a new measure to evaluate of passing as passes are one of the most frequent events in football.

We think that the main weakness of current approaches to evaluate passing, and tactical behaviour in general, is that they seldom include contextual variables and the interaction with the opponent. Above all, these approaches use evaluation strategies based on the probability of goal scoring. This means a pass is classified as a 'good pass' if it produces a goal-scoring opportunity within a set frame of seconds or increases the probability of a shot on goal (Chawla, Estephan, Horton, & Gudmundsson, 2017; Link et al. 2016; Power, Ruiz, Wei, & Lucey, 2017a). Therefore, in most cases, only forward passes are considered good passes, even though a sideways or a backward pass might provide more value. Although these passes might not directly result in scoring opportunities, they might disrupt the defensive organisation of the opponent, creating space for scoring opportunities.

As we highlighted in section 5, disrupting the defensive organisation is a marker for successful ball possession. Based on these assumptions, we believe that tactical performance measures should not be directly related to the probability of goal scoring. Therefore, we developed a new approach to evaluate successful passes (passes reaching a teammate) based spatial aggregates of tracking data. This approach values successful passes by calculating changes in the positioning (organisation) of the defensive team and its subunits following a pass. The rationale behind our approach is that a high amount of player movement and large change in the positioning and organisation of the defence creates a decoupling of the defending and attacking teams. As shown in section 5, this decoupling, or disruption of defensive coupling, characterises successful attacks.

To measure changes in defensive organisation, we use a principle component analysis to merge spatiotemporal aggregates that represent the organisation of a team and its subunits. These aggregates have been previously demonstrated to adequately describe the tactical dynamics of football. To prove the validity of the D-Def score, we show that it is highly connected to the overall movement of the players of the defending team. In addition, we examine whether the D-Def score can differentiate the performance of different passes and players. In the last step, we calculate the predictive values of different passing parameters such as passing velocity, passing length, and passing angle on defensive disruptiveness. This allows us to examine which passes cause high defensive disruptiveness and whether the direction of a pass is an important factor in D-Def score.

CONSTRUCTION OF D-DEF AND I-MOV

To construct our disruption variables, we collected tracking and pass data (16,943 passes) for 18 competitive professional football matches. For our model, we constructed two measures of defensive movement. The first represents the total individual movement of the defensive players (I-Mov) on the field, and the second represents the disruption of the defensive organisation (D-Def).

Our measure of total individual movement is constructed out of two components: absolute displacement (m) in the longitudinal (X-axis) and absolute displacement (m) in the lateral (Y-axis). We first calculated the sum of absolute displacement in the X-position (Xc) and Y-position (Yc) for all defending players between the moment the pass was given (t0) and 3 seconds later (t0 + 3). Second, we concatenated the sums of displacement on the X-axis and Y-axis to construct I-Mov.

$$\mathbf{Xc} = \sum_{(i=1)}^{n} | (X_1^{(t0+3)} - X_1^{t0}) | + |(X_2^{(t0+3)} - X_2^{t0}) | + \ldots + |(X_{11}^{(t0+3)} - X_{11}^{t0}) |$$

$$\mathbf{Yc} = \sum_{(i=1)}^{n} | (Y_1^{(t0+3)} - Y_1^{t0}) | + |(Y_2^{(t0+3)} - Y_2^{t0}) | + \ldots + |(Y_{11}^{(t0+3)} - Y_{11}^{t0}) |$$

$$\text{I-Mov} = Xc + Ycgh$$

To quantify the defensive organisation, we computed the displacement of the average X and Y positions (centroids) for the full team (CX, CY in m), and the defensive (CDEFX, CDEFY in m), midfield (CMIDX, CMIDY in m), and attacking (CATTX, CATTY in m) lines between the moment a pass was given (t0) and three seconds later (t0+3). Line formations were based on the starting formations of the teams as provided by coaches before a match, and we accounted for substitutions. We also computed the change in surface area (SAREA) and the change in team spread (SF) of the full team. All these spatial aggregates are visually explained in Figure 1.

To create an overall composite measure for the disruptiveness of the defensive organisation, we then conducted a principal component analysis (PCA based on these displacement measures of all passes in our data set.

$$PC1 = -0.46C_X + 0.26C_Y - 0.43C_{XDEF} + 0.24C_{YDEF} - 0.43C_{XMID} + 0.24C_{YMID} - 0.41C_{XATT} + 0.24C_{YATT}$$

$$PC2 = -0.26C_X - 0.47C_Y - 0.24C_{XDEF} - 0.43C_{YDEF} - 0.25C_{XMID} - 0.43C_{YMID} - 0.24C_{XATT} - 0.40C_{YATT}$$

$$PC3 = 0.71\, S_{area} + 0.71\, S_F$$

$$\text{D-Def} = |PC1| + |PC2| + |PC3|$$

The resulting D-Def score is a unitless score based on the absolute standardised variables multiplied by the factor loadings. The theoretical range of the D-Def score is 0 to 20. A score of 0 represents no disruption of the defensive organisation at all, whereas a score of 20 would represent the maximal amount of achievable disruption on all components at the same time. In simple terms, the three components of the D-Def can be described as:

- disruption in longitudinal direction (PC 1),
- disruption on lateral direction (PC 2),
- and disruption in overall organisation (PC 3)

EXPERIMENTAL RESULTS

Because individual movement does not necessarily have to result in disruption (e.g. if two players in the same line change position), we wanted to know to what extent an increase in total individual movement after a pass results in a disruption of the defensive organisation. A relatively strong Pearson correlation between I-Mov and D-Def of $R^2 = 0.74$ confirmed our initial hypothesis that both measures are highly connected.

To demonstrate the sensitivity of the I-Mov and D-Def measures, we ranked all 6460 passes in our data set both on the I-Mov measure and on the D-Def measure. As can be seen in Figure 6a, both variables are able to clearly distinguish between top level, average, and bottom level passes validating the sensitivity of them.

We used the same approach to see if the two variables would also be sensible enough to separate players with a high defensive disruption to ones with average and low disruption. Like the pass ranking, I-Mov and D-Def showed a good sensitivity to distinguish between players.

To investigate the characteristics of a disruptive pass, we used corrected One-Way ANOVAs on pass length, pass angle, and pass velocity. Our results showed that more disruptive passes (higher D-Def value) showed longer pass length and higher pass velocity. The passing angle did not influence the disruptiveness, so in this model sideways and backwards passes can also be of high value. Our results revealed that a top-level pass, which leads to a significant decrease in defensive organisation, was between 19 and 30m long and was given with a passing velocity of at least 10.7 m/s.

On an individual level, good passers instigated a lot of defensive movement in the longitudinal direction. Increased forward and backward movements typically results in more space between the defending lines and a decoupling between the subunits, thus creating opportunities for forward passes.

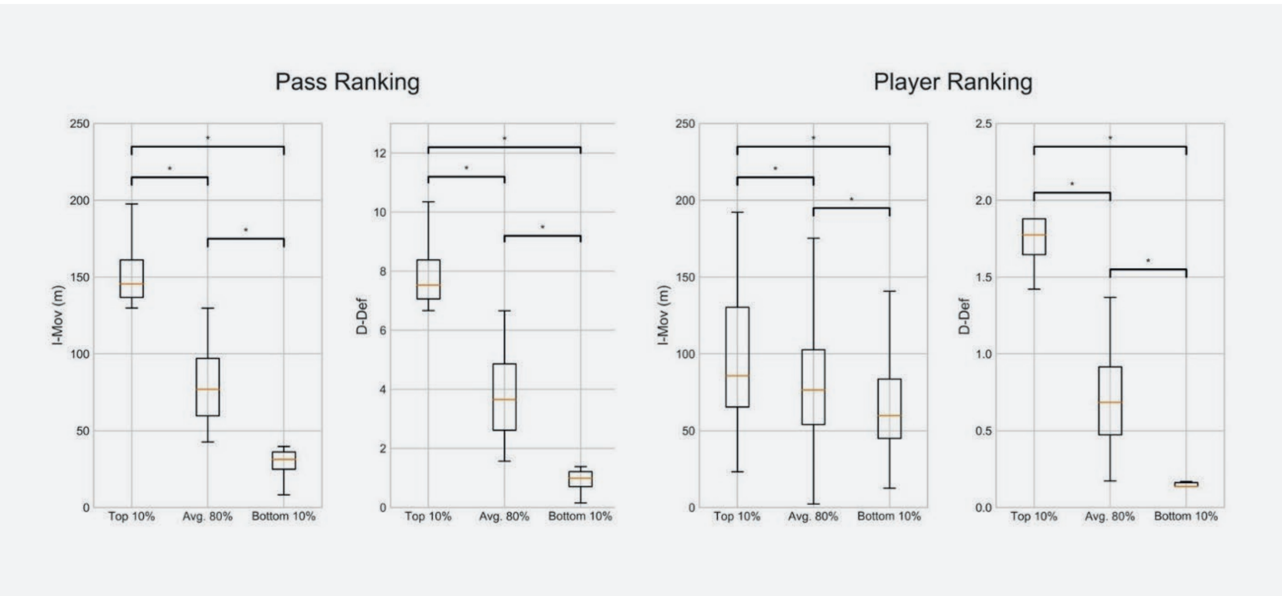


Figure 5. Box Plot passes & players The panel was adapted from work of Goes et al. (2018)

DOES DEFENSIVE DISRUPTION EQUAL EFFECTIVENESS

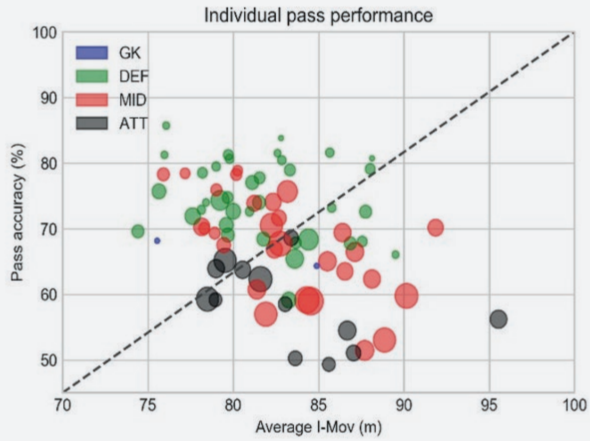
Although the results of our first study establishing the idea of defensive disruption as a measure for performance are quite promising, they left several questions unanswered. One of these questions was: how are the variables associated with playing positions? Given their position on the field, players might have disparate opportunities to execute disruptive passes. A second question was the connection of disruption and passing accuracy, as it makes no sense to label a player as a good passer based on their disruption value if they just complete 20 percent of their passes.

In a second study we tackled these questions using tracking data of 82 professional football matches (Kempe, Goes, & Lemmink, 2018).

Within this study we observed that midfield players, in general, scored highest in defensive disruption. This seems quite natural as they are usually assigned to their position given their superb passing ability. Further, most of their passes occur in the centre of pitch where, theoretically, greater overall movement is possible compared to an attacker who more often plays passes close to the goal with less space for movement to occur.

In addition, we discovered a disruption vs. accuracy trade-off meaning that players with high disruption ratings showed relatively lower passing accuracy values. This indicates that highly disruptive passes are more challenging and riskier. On the other hand, highly disruptive players were also responsible for more key passes and assists indicating that disruption is associated with creating goal scoring opportunities.

	Player	*D-Def*	*PC1*	*PC2*	*PC3*
1	Player 1 (MID)	2.70	−1.63	−0.86	0.10
2	Player 2 (DEF)	1.88	−0.88	0.93	0.08
3	Player 3 (DEF)	1.66	−0.64	0.49	0.53
4	Player 4 (MID)	1.64	0.70	0.40	−0.54
5	Player 5 (DEF)	1.41	0.93	0.29	−0.20
...	...	...	...	...	...
55	Player 30 (MID/ATT)	0.17	0.13	0.02	0.02
56	Player 40 (MID/ATT)	0.16	0.03	0.12	0.01
57	Player 27 (DEF)	0.14	−0.02	0.10	−0.02
58	Player 42 (MID/ATT)	0.13	0.00	0.10	−0.03
59	Player 64 (DEF)	0.08	0.02	−0.05	−0.01



^
Figure 6
On the top, the three components associated with the D-Def score (PC1, PC2, andPC3) of the players are shown. Only the best and worst scoring five players are shown in the table. On the bottom, average pass accuracy vs. average pass effective-ness (I-Mov) per player. Colour repre-sents field position; size of the marker represents the number of key passes per game (more key passes = bigger marker). The panel was adapted from work of Kempe et al. (2019).

IS DEFENSIVE DISRUPTION RELATED TO WINNING?

The next and perhaps most important question for coaches and decision-makers is: can these theory-based variables help win games? Therefore, we investigated the connection of I-Mov and D-Def to winning. To get a better idea if our approach is indeed beneficial, we also calculated other indices that claimed to be associated with winning games.

Within this study (Kempe & Goes, 2019) we updated our previous model in two important ways. First, instead of a three second window, we now normalise the effect of a pass per second. In the previous model we undervalued longer passes as their effect might not be captured in total with the three second window. With this new calculation the effect of a pass is normalised and better represented.

In a second step, we implemented a new way to register team formations which are the basis to calculate the changes in defensive organisation. Therefore, we adapted the idea of Bialkowski et al. (2016) using a K-Nearest Neighbour approach to cluster players in different playing positions and formations. To do so, we used the tracking data of each match to find an attacking and defending formation for both teams. We then assigned every player to one line of the formation (defence, midfield, or attacking) for each second of the game. By applying this idea to our approach, instead of starting formations of a team, we differentiated between offensive and defensive formation and were able to evaluate passes by taking the change of playing positions and formations into account. Both of those updates increase the validity of our approach by reflecting the high amount of variation in the game.

With our updated calculation, we compared the I-Mov and D-Def values of the winners and losers of 89 professional football games. In this analysis, the I-Mov could distinguish between winning and losing teams with a difference in defensive movement of nearly 70%. The effect sizes of the I-Mov and its components were also high, indicating a robust effect.

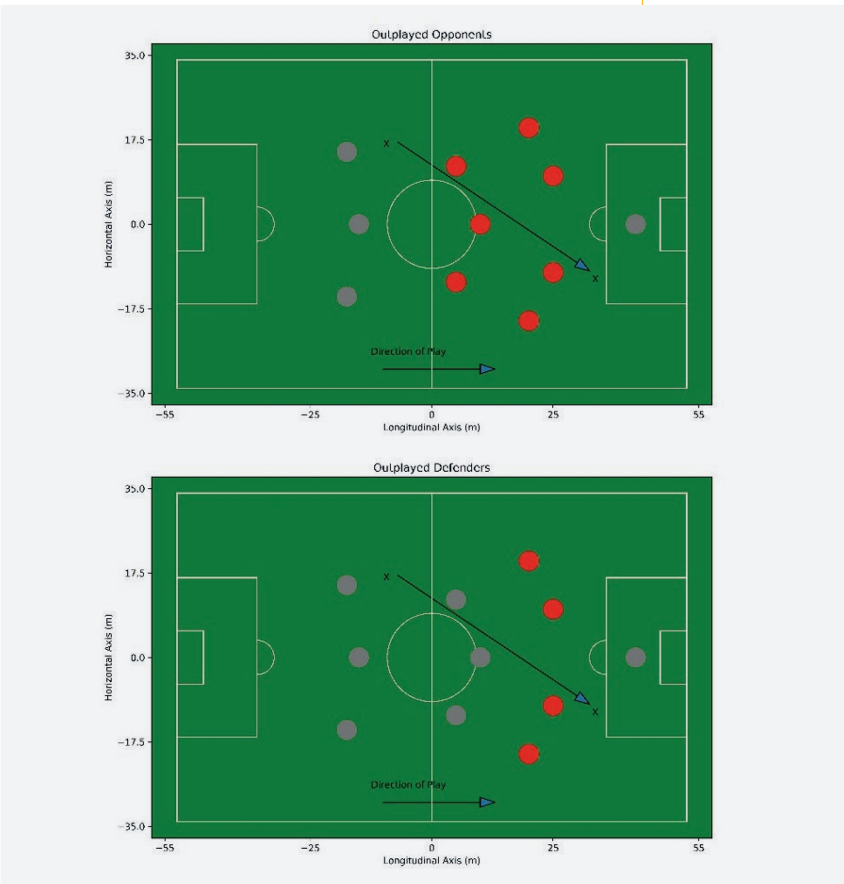
In contrast, the D-Def only differentiated winning and losing teams in the longitudinal (PC 1) and team organisation (PC 2) components, whereas an effect of the overall D-Def was diminished by the lateral component. This indicates that to win a game, a disruption towards the goal and in the overall organisation of the team are more important than disruption side to side.

	WINS (N = 89)	LOSSES (N = 89)	MEAN DIFF.	EFFECT SIZE (COHEN'S D)
INDIVIDUAL MOVEMENT (I-MOV)				
I-Mov-X (Mean)	0.866m ± 0.673m	0.515m ± 0.675m	+68.1%	0.52 ***
I-Mov-Y (Mean)	0.772m ± 0.600m	0.772m ± 0.600m	+71.2%	0.54 ***
I-Mov (Mean)	1.638m ± 1.268m	1.638m ± 1.268m	+69.6%	0.53 ***
DEFENSIVE DISRUPTIVENESS (D-DEF)				
PC1 (Mean	0.018 ± 0.015	0.013 ± 0.022	+34.1%	0.24*
PC2 (Mean)	0.010 ± 0.013	0.014 ± 0.033	-23.6%	-0.13
PC3 (Mean)	-0.026 ± 0.022	-0.021 ± 0.022	-25.5%	-0.25*
D-Def (Mean)	0.474 ± 0.048	0.484 ± 0.072	-2.0%	-0.16

Table 1. Descriptive statistics winning and losing teams (*: p = .05**: p < .05, ***: p < .01)

To compare the disruption approach with other measures of tactical performance we calculated variables for commonly used principles: the zone and the imbalance principle (F. Goes et al. 2019). For the current study we adapted features currently available in science and practice.

To quantify tactical performance on the zone principle, we constructed a simple zone measure that is comparable to expected goals, providing a value from 0 to 1 of the probability to score a goal. The complex zone measure takes ball pressure on the passer and the receiver into account and is partly adapted from the work by Link (Link et al. 2016). Both, simple and complex zones were computed for every successful pass and reception and then aggregated over the full match. To quantify the imbalance principle, we constructed two passing features and three off-the-ball balance features. Our passing features followed the description of the Packing-Rate("Impect," n.d.) and Impect("Impect," n.d.). We computed the number of outplayed opponents based on the longitudinal coordinates of the pass, reception, and all opposing players, and we computed the number of outplayed defenders based on the longitudinal coordinates of the pass, reception and the last six players on the field plus the goalkeeper (Figure 7). Note that the number of outplayed opponents can also be negative in the case of a backwards pass. As off-the-ball balance features, we computed numerical superiority scores for the attacking team on the opposing half, in the final third, and in the penalty area. To do so, we assessed numerical balance (by counting players of both teams) in a certain area during every pass-reception window and awarded points for every window in which the attacking team had numerical superiority in that area (+1 player = 1 point, +2 players = equal 2 points, etc.).



^
Figure 7
Visual representation of outplayed opponents (top) and outplayed defenders (bottom) as a result of a pass. Outplayed opposing players are shown in red, other opposing players are shown in grey.

Based on these calculations, we found that winning teams had a significantly increased mean complex zone score for pass receivers ($H(176) = 4.16$, $p < 0.05$), and a significantly increased superiority score in the final 3rd ($H(176) = 6.90$, $p < 0.01$) and penalty area ($H(176) = 5.09$, $p < 0.05$) compared to losing teams. However, compared to the defensive disruption measures, these variables show relatively low effect sizes. These results indicate that some of the features that have gained considerable popularity within the analytics community over the recent years seem to have limited practical value.

In summary, defensive disruptiveness variables reliably distinguish between winning and losing teams. They also outperform other commonly used tactical performance variables, which highlights the usefulness of this approach.

	WINS (N = 89)	LOSSES (N = 89)	MEAN DIFF.	EFFECT SIZE (COHEN'S D)
		ZONE PRINCIPLE		
Simple zone passer (Mean)	0.031 ± 0.013	0.028 ± 0.012	+10.7%	0.24
Simple zone receiver (Mean)	0.040 ± 0.014	0.037 ± 0.014	+8.1%	0.24
Complex zone passer (Mean)	0.022 ± 0.010	0.020 ± 0.010	+10%	0.21
Complex zone receiver (Mean)	0.032 ± 0.012	0.028 ± 0.011	+14.3%	0.28 *
		BALANCE PRINCIPLE		
Outplayed defenders (Mean)	0.23 ± 0.10	0.21 ± 0.09	+9.5%	0.19
Outplayed opponents (Mean)	0.39 ± 0.17	0.39 ± 0.16	+2.6%	0.13
Half Superiority (Total)	2.82 ± 7.67	1.87 ± 5.78	+50.8%	0.14
Final 3rd Superiority (Total)	3.11 ± 3.52	2.22 ± 3.04	+40.0%	0.27 **
Score Box Superiority (Total)	0.84 ± 1.51	0.76 ± 3.39	+10.5%	0.03 *

Table 3. Descriptive statistics (mean ± std.) of winning and losing teams on the various principles of play. * (p < .05) and ** (p < .01) denote significant differences between winning and losing teams.

As a next step, we wanted to see if we could not just can differentiate between winners and losers, but predict match outcome based on the mean total movement feature (I-MovMean, as it captures both the movement in longitudinal as well as lateral direction), mean longitudinal disruption feature (PC1Mean), and mean surface disruption feature (PC3Mean). We chose this combination of features based on their discriminative power and the fact that the combination of these features yielded the highest accuracy and lowest log loss scores. To do so, we first split the data set in a training set that contained 80% of the data, and a test set that contained 20% of the data, stratified on match outcome. Furthermore, we scaled and fitted a 5-fold cross-validated Logistic Regression model to our training dataset and predicted winning and losing probability for both teams in every match. Based only on these three simple components we were able to predict binary match outcome with an accuracy of 69.4% and a log loss of 0.65, based on the following regression equation (3):

$$\textbf{Outcome} = -0.146 + 0.689\ \text{I-Mov}_{Mean} + 0.172\ \text{PC1}_{Mean} - 0.592\ \text{PC3}_{Mean}$$

CONCLUSION

In conclusion, with the D-Def we have developed a new measure to quantify passing effectiveness, a major part of tactical performance. We accomplished this by building on years of research in sports and behavioural science using the principles of the complex (dynamic) systems approach, and combined it with basic data-science techniques within a multi-dimensional team. Although the D-Def itself can still be improved, it highlights that in order to evaluate tactical performance it is important to study the behaviour and interaction of teams instead of simple event based and goal-centric statistics.

REFERENCES

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T., & Weber, H. (2017). *Visual Analysis of Pressure in Football. Data Mining and Knowledge Discovery* (Vol. 31). Springer US. https://doi.org/10.1007/s10618-017-0513-2

Balague, N., Torrents, C., Hristovski, R., Davids, K., & Araújo, D. (2013). Overview of Complex Systems in Sport. *Journal of Systems Science and Complexity*, 26(1), 4–13. https://doi.org/10.1007/s11424-013-2285-0

Bartlett, R., Button, C., Robins, M., Dutt-Mazumder, A., & Kennedy, G. (2012). Analysing Team Coordination Patterns from Player Movement Trajectories in Soccer: Methodological Considerations. *International Journal of Performance Analysis in Sport*, 12(2), 398–424. https://doi.org/10.1080/24748668.2012.11868607

Bialkowski, A., Lucey, P., Carr, P., Matthews, I., Sridharan, S., & Fookes, C. (2016). Discovering Team Structures in Soccer from Spatiotemporal Data. *IEEE Transactions on Knowledge and Data Engineering*, 28(10), 2596–2605. https://doi.org/10.1109/TKDE.2016.2581158

Brink, M. S., Kuyvenhoven, J. P., & Toering, T. (2018). What Do Football Coaches Want from Science?. *Kinesiology*, 50, 150-154

Chawla, S., Estephan, J., Horton, M., & Gudmundsson, J. (2017). Classification of Passes in Football Matches using Spatiotemporal Data. *ACM Transactions on Spatial Algorithms and Systems*, 3(2), 6–36. https://doi.org/10.1145/3105576

Davids, K., Araújo, D., & Shuttleworth, R. (2005). Applications of Dynamical Systems Theory to Football. *Science and Football V*, (i), 537–550.

Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and Inter-Group Coordination Patterns Reveal Collective Behaviors of Football Players Near The Scoring Zone. *Human Movement Science*, 31(6), 1639–1651. https://doi.org/10.1016/j.humov.2012.03.001

Fernandez, J., & Bornn, L. (2018). Wide Open Spaces : A Statistical Technique for Measuring Space Creation in Professional Soccer. MIT Sloan Sports Analytics Conference, (March), 1–19.

Frencken, W., Lemmink, K., Delleman, N., & Visscher, C. (2011). Oscillations of Centroid Position and Surface Area of Soccer Teams in Small-Sided Games. *European Journal of Sport Science*, 11(4), 215–223. https://doi.org/10.1080/17461391.2010.499967

Goes, F., Kempe, M., Lemmink, K., Goes, F., Kempe, M., & Lemmink, K. (2019). Predicting Match Outcome in Professional Dutch Football Using Tactical Performance Metrics Computed from Position Tracking Data

Goes, F. R., Meerhoff, L. A., Bueno, M. J. O., Rodrigues, D. M., Moura, F. A., Brink, M. S., ... Lemmink, K. A. P. M. (2020). Unlocking the Potential of Big Data to Support Tactical Performance Analysis in Professional Soccer: A Systematic Review. *European Journal of Sport Science*. https://doi.org/10.1080/17461391.2020.1747552

Goes, Floris R., Kempe, M., Meerhoff, L. A., & Lemmink, K. A. P. M. (2018). Not Every Pass Can Be an Assist: A Data-Driven Model to Measure Pass Effectiveness in Professional Soccer Matches. *Big Data*, 6(4). https://doi.org/10.1089/big.2018.0067

Gréhaigne, J.-F., Godbout, P., & Bouthier, D. (1999). The Foundations of Tactics and Strategy in Team Sports. *Journal of Teaching in Physical Education*, 18(2), 159–174. https://doi.org/10.1123/jtpe.18.2.159

Gréhaigne, J. F., Bouthier, D., & David, B. (1997). Dynamic-system Analysis of Opponent Relationships in Collective Actions in Soccer. *Journal of Sports Sciences*, 15(2), 137–149. https://doi.org/10.1080/026404197367416

Grosskreutz, H., & Rüping, S. (2009). On Subgroup Discovery in Numerical Domains. *Data Mining and Knowledge Discovery*, 19(2), 210–226. https://doi.org/10.1007/s10618-009-0136-3

Gudmundsson, J., & Horton, M. (2017). Spatio-Temporal Analysis of Team Sports. *ACM Comput*. Surv., 50(2), 22:1--22:34. https://doi.org/10.1145/3054132

Haken, H., Kelso, J. A. S., & Bunz, H. (1985). A Theoretical Model of Phase Transitions in Human Hand Movements. *Biological Cybernetics*. https://doi.org/10.1007/BF00336922

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C., & Meyer, T. (2019). Machine Learning in Men's Professional Football: Current Applications and Future Directions for Improving Attacking Play. *International Journal of Sports Science and Coaching*, 14(6), 798–817. https://doi.org/10.1177/1747954119879350

Impect. (n.d.).

Kempe, M., Goes, F., Kempe, M., & Goes, F. (2019). Move It or Lose It : Exploring the Relation of Defensive Disruptiveness and Team Success

Kempe, M., Goes, F. R., & Lemmink, K. A. P. M. (2018). Smart Data Scouting in Professional Soccer: Evaluating Passing Performance Based on Position Tracking Data. *Proceedings – IEEE 14th International Conference on eScience, e-Science 2018*. https://doi.org/10.1109/eScience.2018.00126

Link, D., Lang, S., & Seidenschwarz, P. (2016). Real Time Quantification of Dangerousity in Football Using Spatiotemporal Tracking Data. *PLoS ONE*, 11(12), e0168768. https://doi.org/10.1371/journal.pone.0168768

Power, P., Ruiz, H., Wei, X., & Lucey, P. (2017a). "Not All Passes Are Created Equal:" Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Vol. Part F1296, pp. 1605–1613). New York, NY, USA: ACM. https://doi.org/10.1145/3097983.3098051

Rein, R., & Memmert, D. (2016, December 24). Big Data and Tactical Analysis in Elite Soccer: Future Challenges and Opportunities for Sports Science. *SpringerPlus*. https://doi.org/10.1186/s40064-016-3108-2

Rein, R., Raabe, D., & Memmert, D. (2017). "Which Pass Is Better?" Novel Approaches to Assess Passing Effectiveness in Elite Soccer. *Human Movement Science*, 55, 172–181. https://doi.org/10.1016/j.humov.2017.07.010

Siegle, M., & Lames, M. (2013). Modeling Soccer by Means of Relative Phase. *Journal of Systems Science and Complexity*, 26(1), 14–20. https://doi.org/10.1007/s11424-013-2283-2

Spearman, W., Basye, A., Dick, G., Hotovy, R., & Pop, P. (2017). Physics-Based Modeling of Pass Probabilities in Soccer, 1–14.

Reorganising around the ball in locational play: dynamic and multi-directional rebalancing

— Isaac Guerrero[1], Xavier Damunt[1], Marc Quintana[1], Jordi Fernández[1] & Paco Seirul·lo[1]

1: FC Barcelona

INTRODUCTION

Upon analysing football both at the macro-level of competition and the micro-level of an individual match, it becomes clear that unpredictability and dynamism are the defining characteristics of the sport (Couceiro et al. 2014; Ric et al. 2016).

Football is based on a series of constantly evolving and very specific perceptual and motor challenges. Given this observable reality, there is a clear answer as to what the foundation of our methodology should be: firstly, the decision-making system should be optimised in order to move beyond traditional PAD + E (Perception – Analysis – Decision + Execution) and make room for concepts such as feeling and emotional emergence (Damunt & Guerrero, 2018; Damunt & Guerrero, 2020). Secondly, the team's dynamic balance must be enhanced to prevent disorganisation and remain efficient regardless of the specific events and issues that take place during a match. Of all the sports we have observed and analysed, football ranks the highest in uncertainty. This discipline stands apart largely due to two factors:

- The moving object – the ball – is manoeuvred mainly by a non-dominant limb, or rather, the legs. This point alone clearly indicates just how complicated it may be for players to take decisions in advance, at least in close proximity to the ball.
- The same non-dominant limb responsible for executing the motor action is, at the same time, also in charge of moving the body through space and, thus, shifting its axis (Seirul·lo, 1981).

The foundation of football coaching and tactics has traditionally been the coordination of players' movements along the horizontal axis of the field (Duarte et al. 2013) and towards the goals (Silva et al. 2016). Within this view of the sport, the relationships between the lines of play that mark conventional positions of the field (goalkeeper, defenders, midfielders, forwards) represent a focal point (Gonçalves et al. 2014). Stemming from the concept of locational play, which has a trajectory of more than 40 years in FC Barcelona, our methodological proposal responds to scientific findings on complexity in sports that have represented a decisive step towards a paradigm shift in the practice and optimisation thereof.

Aware of the random and stochastic nature of the game and the need for players to self organise, Johan Cruyff designed and optimised this methodology during the time he spent at FC Barcelona alongside Paco Seirul·lo. Coaches Pep Guardiola and Tito Vilanova, as well as technical coordinators such as Joan Vilà, later worked to refine it (Buldú et al. 2019; Gyarmati et al. 2014; Herrera-Diestra et al. 2020).

The difference in our understanding of the sport lies in the fact that we self-organise around the moving object, the ball. In doing so, we treat other references in the game as relative, including goals, lines, zones, boundaries, the opponent, etc. This organisation around the ball, expressed through concepts such as Redistribution, Relocation and Recovery (3Rs), allows the team to reorganise with a single collective intention in mind: to be in possession of the ball for as long as possible (Lago et al. 2007) in order to break down the opponent's symmetries, disorganise them, and recover the ball as quickly as possible upon losing possession.

Viewing the development of the match from this perspective of circular organisation also diminishes the importance of so-called "progression". At the same time, it gives rise to the ongoing search for advantageous spaces of phase (see next section) – understood to be the continuum of everything taking place on the playing field at any given time – and constrains certain potential actions that players might take. These actions can be observed through the shaping elements of our game idea that define the athlete's location based on her/his

situation and position. We're referring specifically to gesture, profile and orientation in relation to space and time (Arbues-Sanguesa et al. 2020). Together with the player's present and future intentions, these aspects help determine whether or not an advantage has been gained in the match. Take, for example, a situation in which three nearby players pass the ball amongst themselves without any apparent intent to advance. At first glance, this may seem to be both a disadvantage and a waste of the space available. In the proposed methodology, however, what we actually observe is that this chain of seemingly fruitless and safe interactions has the potential to lead to a burst of high-level entropy in our organisation, thus disorganising the opponent. Hristovski and Balagué (2020) recently described a series of principles based on the concept of entropy in cooperative and competitive environments. Being in disposition of the ball should yield more passing opportunities and offer more information about the environment (opponents) through the establishment of a dynamic balance. Here, possession does not respond to a predefined plan or a top-bottom organisation but the ability to use the available information of the environment (opponents) to adapt and reduce the likelihood of losing the ball. Therefore, this type of organisation will be more unpredictable and, thus, more efficient, increasing the potential courses of action. From a comprehensive view of the game, this is explained through the concepts of self-organisation, a fundamental principle of complex systems and which acts on all levels (from the micro to the macro and from the individual to the collective) (Araujo et al. 2015). Subject to the regulations and the resulting logic of the game, both teams participate with the aim of achieving their goal while simultaneously working to prevent the opponent from achieving theirs. This interplay of intentions is what leads each team to self-organise, with opposing objectives, and to display a symmetrical relationship in their interactions, which can make the match appear stable at times. An advantage is gained when one of the teams manages to destabilise the other, breaking this symmetry and imposing their intentions on their opponent. This generates a sense of instability in the match, as well as new space and time conditions that benefit the team in question.

Because the reference in this approach, that is, the ball, is neither static (as are the lines, goals, zones, etc. used in other ideas) nor proactive or reactive (as the opponents may be) but rather dynamic, it is possible to:

- Maintain a continuous dynamic balance in the distance between players and between players and the ball. To do so, efficient locational dynamism is required from every member of the team, since it promotes the multi-directional distribution of the ball.
- Be available around the ball, with some players further away and others closer, forming two contour curves around it. This is the case both when in possession of the ball and not.
- Communicate amongst ourselves through assertive motor skills, as well as contra-communicating against the other team (Parlebas, 2001).

CONSTANT DYNAMIC BALANCE

This entire approach stems from a different interpretation of space and time, one based on a stochastic sequence of spaces of phase that ensure continuity in the game through the intentions of our players. These intentions are expressed through Redistribution, Relocation and Recovery (3Rs), which occur as the players seek out a dynamic rebalancing primarily around the ball, but also around their teammates and opponents. This all takes place under the very strong influence of the regulations. Football is governed by a series of rules, such as offside and unlimited possession time, which invite us to interpret the structural cohesion and the underlying collective dynamic of the players differently than we might in most other sports. Unlike other disciplines, including invasion sports such as rugby, football players can generally act multi-directionally. The concept of locational play, developed by FC Barcelona, makes the most of this special and remarkable particularity, made possible by a set of regulations that has undergone very few changes in the past 150 years and none at all which significantly impact the freedom of direction. The shaping elements of this "natural" interpretation of the game (Seirul·lo et al. 2019) are based on the special features of the sport, among other things. An example would be the possibility of interacting during the match in our own half, unlike other sports in which this type of interaction is restricted, such as futsal or basketball. Other special features of the sport include unlimited possession time or the lack of any rules that incites progression, such as the forward pass in rugby. The vast majority of playing styles today have a line-based organisation that uses width and depth as the preferred reference rather than proximity to the ball to offer as many communication channels (passing lines) as possible. The other styles of play show a pronounced tendency toward progression and disregard the regulatory advantages available. In contrast, our players make the most of them and interact in multiple directions at 360°.

Therefore, we believe that organisation should be based chiefly on the location of the ball, which we understand as the centre of the game, and not on the occupation of pre established spaces. The idea is to redirect the focus and view the game from a different, not so "tactical" but rather communicative perspective. Players reorganise around the ball, constantly (motor) communicating with each other (Parlebas, 2001) through the moving object. This allows them to continuously establish optimal distances and communication channels with their teammates and make interpretations based on the ball's trajectory. We understand the game through the creation of spaces of phase in multiple directions. The aim is to promote players to constantly relocate themselves around the ball: every time the ball changes location, the players will reorganise around that location, occupying a unique space and time at a determined speed and pace. Changing locations in this way results in the creation of a significant number of communication channels at any given moment, which makes it possible to redistribute the ball in multiple directions. Hypothetically, these multiple possible channels should generate some amount of disorganisation in the opponent, allowing the channel toward the goal, instead of a teammate, to emerge as the clearer option when the occasion arises.

This type of collective self-organisation has proven to be more efficient than any other line-based interpretation of the structural dynamics of football. According to a case study published by Buldú, Busquets, Echegoyen and Seirul·lo in Scientific Reports (2019) that looked at FC Barcelona's first team, coached by Pep Guardiola, we can conclude – with the help of network sciences – that teams that use this type of dynamic reorganisation tend to be more robust and have a greater propensity for passing formations that involve more than three players, as compared to other teams that apply different collective relationships. There is also greater inter-team interaction, which leads to better structural cohesion. Moreover, this may allow for more short passing formations (more passes), which, in turn, would allow our team to manage the playing space comparatively better than their opponents since the ball can be passed from more locations.

FORMING TWO CONTOUR CURVES AROUND THE PLAYER WITH THE BALL

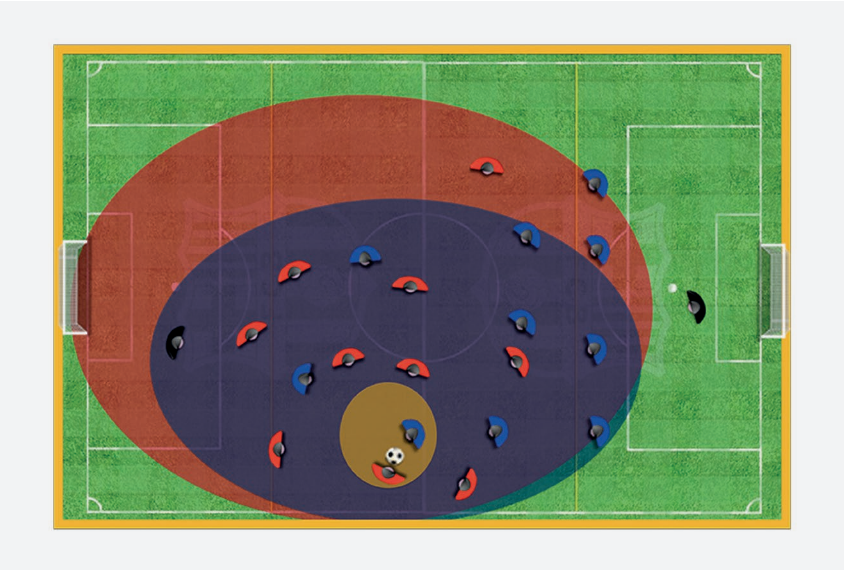
Ideally, organising around the ball means that all players must continuously modify their location based on the ball movement. While constantly relocating, players take into account the location of their teammates and opponents, among other information, in order to establish optimal distances to communicate (and contra-communicate) through the ball and via some basic skills: playing the way you are facing, pass to feet, one-touch pass, and dynamic pass.

Due to the constant relocating, players must interpret their role and that of their teammates at all times. The contour curves or curve levels are the main reference for doing so. When possible given the playing field and offside area, there should be players in the role of mutual support located at 360° around the reference point (the ball and the player with it), both when the team has possession and when it does not. Further away, but still relatively close, are other players in the role of cooperation (Buldú et al. 2019). Both roles are intrinsically related to the socio-affective spaces defined by Seirul·lo (2004) with regard to the space and time available, which define the set of potential actions for the players and team, together with other contextual parameters.

The constant movement of the players and the ball means that roles are occasionally exchanged, and teammates might move from a location of mutual help to one of cooperation and vice versa. Likewise, when the team has possession of the ball, players move into the role of intervention when they receive it. These roles complement one another depending on whether the players are in front of or behind the ball and outside or inside the playing field; traditional positions, thus, lose their significance. Consequently, players are not specialists in only one specific role. Rather, they are able to act based on the moment at hand, in whatever socio-affective space they find themselves in.

If we froze time and took an aerial shot of the space of phase in a match, we could trace one contour curve around mutual support players and another around those in a cooperation role. We would observe two isolines joining the socio-affective points. This organisation exists within the static limits of the playing field, which, in turn, is subdivided into four corridors and four horizontal sections, in addition to the moving offside reference. Unlike other types of organisation, locating players around the ball with contour curves that are ideal for practising locational play means abandoning more distant corridors and zones in order to achieve as many usable or unusable communication channels for the intervening player as possible, depending on whether he is a team-mate or opponent.

Figure 1.
Visual representation of the contour curves.
v




FCB
5
RESPECT
UEFA

COMMUNICATING THROUGH ASSERTIVE-MOTOR SKILLS AMONGST OURSELVES AND CONTRA-COMMUNICATING AGAINST THE OTHER TEAM

Motor communication is transmitted intrinsically when a segmental movement is initiated, for example, with the lower limbs, by modifying the vertical axis and/or moving towards the remote areas (Seirul·lo, 1981) where the sending player can direct his action thanks to an access channel. Because our sport is practised in such reduced space and time conditions, the communication between teammates must provide fluid and clear information. Therefore, the constant location changes resulting from each player's intentions must be recognisable for their teammates. These intentions must be shared. Several factors can help with this, some related to the organisation around the ball and others that must be adopted by the players. In the first case, there is the fact that all the players are familiar with the particularities of each role thanks to constantly exchanging them. This is something the coach and facilitator (Williams & Wigmore, 2020) can examine through the practice sessions they design for players throughout the week. Having this knowledge means it is possible to empathically intuit a teammate's needs. When this capacity is fully optimised, empathic resonance between the players occurs; the players recognise their needs merely from how the motor action is initiated and can anticipate the execution thereof in order to respond to these needs (Damunt & Guerrero, 2020). This fact highlights the importance of bringing together like-minded, non-complementary players. Doing so will make for much more fluent communication comprehensible to all since the players will interpret the game in the same way (Bransen & Van Haaren, 2020).

Players should feel the need to participate in the game in consonance with the intentions of their teammates, or rather, of the team, and with the club's values (Balagué, et al. 2019). Assertive-motor communication is only possible if team members understand each other's intentions, which is easier when they share the same intentions (Silva et al. 2013). If, for example, a player changes his regular locational-play behaviour due to an adverse outcome and, ceasing to use the aforementioned basic skills, attempts to overcome opponents in the direction of the opponent's goal, his teammates may not understand his intentions. This would make it more difficult for him to later relocate himself and maintain the communication channels used in the team's playing style.

Therefore, one of the main functions of

the coach and facilitator is to prevent this type of aggressive-motor behaviour (Guerrero & Damunt, 2016) resulting from the interaction of contextual, and thus temporal, constraints, as well as the intrinsic constraints of the player. The challenge facing the coaches is to integrate these players into the team style of play style while respecting their intrinsic dynamic (Davids, Button, & Bennett, 1999) and natural behaviours (Guerrero & Damunt, 2019). In other words, blend together the natural interaction of the players without losing the essence of the multi-directional play that arises from organising around the ball and without restricting their creativity but rather enhancing it for the common good.

CONCLUSION

“Natural play”, as Paco Seirul·lo put it, takes us back to the essence of the game and the playing style of each and every one of us. To respect the game’s nature, it is first necessary to integrate the intentions of the team members, which must be communicated through assertive and motor skills. Organisation around the ball also makes another natural aspect of the game possible: the need for players to be in constant contact with the ball. Moreover, if the players act (communicate) using their basic skills and in multiple directions, that natural playing style will emerge: locational play. The idea is to get back to the matches we used to play in the town park, during recess, on the streets (Machado et al. 2019): contexts in which we did not think about reducing zones, maintaining the line, offering defensive cover, the ability to keep a shape, etc. Rather, we thought about going for the ball and playing around it so we could pass it around.

> **“The best future path for football would be one heading back to its roots”**

ACKNOWLEDGEMENTS

We would like to thank our players. Every day, they teach us about the special nature of this sport.

REFERENCES

Araújo, D., Passos, P., Esteves, P., Duarte, R., Lopes, J., Hristovski, R., & Davids, K. (2015). The Micro-Macro Link in Understanding Sport Tactical Behaviours: Integrating Information and Action at Different Levels of System Analysis in Sport. *Movement & Sport Sciences-Science & Motricité*, (89), 53-63. https://doi.org/10.1051/sm/2015028

Arbues-Sanguesa, A., Martin, A., Fernández, J., Ballester, C., & Haro, G. (2020). Using Player's Body-Orientation to Model Pass Feasibility in Soccer [Paper presentation]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 886-887). https://openaccess.thecvf.com/content_CVPRW_2020/papers/w53/Arbues-Sanguesa_Using_Players_Body-Orientation_to_Model_Pass_Feasibility_in_Soccer_CVPRW_2020_paper.pdf

Balagué, N., Pol, R., Torrents, C. et al. (2019). On the Relatedness and Nestedness of Constraints. *Sports Med - Open*, 5, 6. https://doi.org/10.1186/s40798-019-0178-z

Bransen, L., & Van Haaren, J. (2020, March 6-7). Player Chemistry: Striving for a Perfectly Balanced Soccer Team [Paper presentation]. 14th annual MIT Sloan Sport Analytics Conference, Boston, USA. http://www.sloansportsconference.com/wp-content/uploads/2020/02/Bransen_paper_player_chemistry.pdf

Buldú, J.M., Busquets, J., Echegoyen, I. et al. (2019). Defining a Historic Football Team: Using Network Science to Analyze Guardiola's F.C. Barcelona. *Sci Rep* 9, 13602. https://doi.org/10.1038/s41598-019-49969-2

Castellano, J., & Pic, M. (2019). Identification and Preference of Game Styles in Laliga Associated with Match Outcomes. *International Journal of Environmental Research and Public Health*, 16(24), 5090. https://doi.org/10.3390/ijerph16245090

Couceiro, M. S., Clemente, F. M., Martins, F. M., & Machado, J. A. T. (2014). Dynamical Stability and Predictability of Football Players: The Study of One Match. *Entropy*, 16(2), 645-674. https://doi.org/10.3390/e16020645

Damunt, X. & Guerrero, I. (2018). The Somatic Markers as a ToolFor Optimization of the Decision-Making System of the Football Player. *United Soccer Coaches Journal*.

Damunt, X. & Guerrero, I. (2020). *El entrenamiento sistémico basado en las emociones. Propuestas para la optimización del jugadoren el fútbol formativo*, Barcelona, FDL.

Davids, K., Button, C., & Bennett, S. (1999). Modeling Human Motor Systems in Nonlinear Dynamics: Intentionality and Discrete Movement Behaviors. *Nonlinear Dynamics, Psychology, and Life Sciences*, 3(1), 3-30. https://doi.org/10.1023/A:1022429522099

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P., & Richardson, M. J. (2013). Competing Together: Assessing the Dynamics of Team–Team and Player–Team Synchrony in Professional Association Football. *Human movement science*, 32(4), 555-566. https://doi.org/10.1016/j.humov.2013.01.011

Guerrero, I, & Damunt, X. (2016, Decembre 3). [5o día del entrenador catalán]. Barcelona, Spain.

Guerrero, I., & Damunt, X. (2019). Methodological, Pedagogical & Didactic Exchange with the Player. Optimization of the Decision-Making System in Youth Football. *Actividad física y deporte: ciencia y profesión*, 31, 19-22. http://colefcafecv.com/wp-content/uploads/2019/07/Revista-COLEFCV-31.pdf

Guerrero, I., Damunt, X. & López, J. (2017). The Creation of a Non-Verbal Communication Code. Working a Proposal for Improving Decision Making in Youth Football. *Actividad física y deporte: ciencia y profesión*, 26, 21-22. http://colefcafecv.com/wp-content/uploads/2017/06/Revista-COLEF-CV-n26.pdf

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). Searching for a Unique Style in Soccer [Paper presentation]. KDD Workshop on Large-Scale Sports Analytics. https://arxiv.org/abs/1409.0308

Herrera-Diestra, J. L., Echegoyen, I., Martínez, J. H., Garrido, D., Busquets, J., Io, F. S., & Buldú, J. M. (2020). Pitch Networks Reveal Organizational and Spatial Patterns of Guardiola's FC Barcelona. *Chaos, Solitons & Fractals*, 138, 109934. https://doi.org/10.1016/j.chaos.2020.109934

Hristovski, R. & Balagué, N. (2020). Theory of Cooperative Competitive Intelligence: Principles, Research Directions and Applications. *Frontiers in Psychology*, 11, 2220. https://doi.org/10.3389/fpsyg.2020.02220

Lago Peñas, C., Martín Acero, R., & Seirul·lo Vargas, F. (2007). Team Performance in Soccer. Modelling the Main Variables Of F.C. Barcelona. *Apunts. Educación Física y Deportes*, 90, 51-58. https://www.revista-apunts.com/apunts/articulos//90/es/090_051-058ES.pdf

Machado, J. C., Barreira, D., Galatti, L., Chow, J. Y., Garganta, J., & Scaglia, A. J. (2019). Enhancing Learning in the Context of Street Football: A Case for Nonlinear Pedagogy. *Physical Education and Sport Pedagogy*, 24(2), 176-189. https://doi.org/10.1080/17408989.2018.1552674

Parlebas, P. (2001). *Léxico de praxeología motriz*, Barcelona, Paidotribo.

Ric, A., Torrents, C., Gonçalves, B., Sampaio, J., & Hristovski, R. (2016). Soft-Assembled Multilevel Dynamics of Tactical Behaviors in Soccer. *Frontiers in psychology*, 7, 1513. https://doi.org/10.3389/fpsyg.2016.01513

Seirul·lo, F. (2004). Estructura Socio-Afectiva. *Documento INEFC Barcelona. Universidad de Barcelona.*

Seirul·lo, F. (1981). Espacio-Tiempo en la Educación Física de Base. *Apuntes de Medicina Deportiva* 18 (70). https://www.apunts.org/en-pdf-X0213371781047438

Silva, P., Chung, D., Carvalho, T., Cardoso, T., Davids, K., Araújo, D., & Garganta, J. (2016). Practice Effects on Intra-Team Synergies in Football Teams. *Human movement science*, 46, 39-51. https://doi.org/10.1016/j.humov.2015.11.017

Silva, P., Garganta, J., Araújo, D., Davids, K., & Aguiar, P. (2013). Shared Knowledge or Shared Affordances? Insights from an Ecological Dynamics Approach to Team Coordination In Sports. *Sports Medicine*, 43(9), 765-772. https://doi.org/10.1007/s40279-013-0070-9

Williams, A. M. & Wigmore, T. (2020) *The Best: How Elite Athletes Are Made*, Australia, Hachatte

8

Ball positioning as key to understanding collective behaviour

— Maurici A. López-Felip[1,2], Matt N. Petersen[1],
Henry S. Harrison[3] & James A. Dixon[1]

1: Center for the Ecological Study of Perception and Action, Department of Psychological Sciences, University of Connecticut, United States

2: FC Barcelona

2: Kognia Sports Intelligence

INTRODUCTION

What is the appropriate approach to understanding football scientifically? Any instance of a game of football is a phenomenon of nature. The events taking place therein include interactions of various kinds, e.g. physical, physiological, cognitive, emotional, social, even political. Clearly, an approach that would divide the analysis of football into different branches of science would face serious limitations due to the incommensurability of the models proposed in each discipline (e.g. sociology, psychology and kinesiology). Due to this incommensurability, the phenomenon is mistakenly taken to be sui generis in every domain.

To address these issues in football (aka soccer) from a multidisciplinary perspective, the Center for the Ecological Study of Perception and Action (CESPA) at the University of Connecticut and the Barça Innovation Hub at Futbol Club Barcelona (FCB) entered into a joint partnership in 2017. The objective of this partnership is to pursue a principled scientific perspective on the game of football, treating its subject matter as rigorously as astrophysicists treat the formation of galaxies or quantum physicists treat the study of elementary particles.

CESPA is a research centre at the University of Connecticut dedicated to uncovering the general principles that govern the behaviour of living systems. Following in the tradition of ecological psychology (Gibson, 1979), CESPA seeks to explain behaviour with respect to the meaningful components of an organism's surroundings at its own scale, the ecological scale, rather than in terms of the physical rules of a smaller scale. The scientific laws that govern the behavioural scale, ecological physics, are pursued by research in areas such as complex systems, dynamical systems theory, and non-equilibrium thermodynamics. In the past, CESPA has participated in collaborations ranging from biomechanics (see e.g. Pagano & Turvey, 1998; Turvey & Fonseca, 2014) and sports science (see e.g. López-Felip, Davis, Frank & Dixon, 2018) to statistical physics and self-organizing systems (Shaw & Kinsella-Shaw, 1988; Frank, 2015; Dixon, Kondepudi, Kay & Davis, 2016).

Now, together with FCB, we aim to turn these methods to football and develop an understanding of the game as a natural phenomenon through the lenses of theory, modeling, and empirical research.

Addressing these issues implies finding the right ontology. The CESPA group, inspired by the work of James Gibson, contends that a fundamental misconception of traditional scientific approaches to behaviour is that of organism-environment dualism. This dualism treats the agent and its world apart as separate spheres abiding by different principles (the environment governed by physics, the organism by psychology). Instead, the ecological approach takes the organism-environment system as a whole as its unit of study.

Our partners at FCB, inspired by the coaching philosophy of Johann Cruyff, complement this perspective in their contention that traditional perspectives on football are similarly mistaken. Specifically, the game of football has traditionally been divided in two separate phases and groups of actions: attack (by the team controlling the ball) and defence (by their opponent). A long-term culture of the game within the Catalan football club, however, rejects this offense-defense dualism, arguing that this perspective unnecessarily and artificially divides the game into two ontological entities disconnected from football as it is actually played. Instead, the FCB philosophy proposes a new ontology of the game, one in which comprehension of the game is centred on the ball.

These simple rules and objectives of football create the constraints and conditions leading to the organisation of behaviour during gameplay. From the young novice who always chases the ball, to the more advanced tactical patterns of a seasoned expert, a player's behaviour is always determined by the constraints of the game which in turn are created by the ball. In other words, the ball's position constitutes the fundamental predictors for the behaviour of the players. Therefore, any model of football dynamics, regardless of which aspect of the game or the level of analysis it targets, must contain parameters describing the position of the ball. Specifically, how the players' behaviours change as a function of these parameters is the natural follow-up question.

Following these considerations, the approach of CESPA and FCB is to combine the different interactions of football (physical, physiological, cognitive, emotional, social, etc.) as aspects of a complex system. The idea is to apply the tools of an extended physics capable of dealing with circular causality and end-directed behaviour (Rosen 1991). This is not a reductionism that would explain, for example, the decision to shoot in terms of neuronal action potentials. Instead, we seek natural laws at the ecological scale, in which physical laws might be commensurate with behaviour (Turvey & Shaw, 1995). This alternative, coined by Gibson as ecological physics (Gibson, 1979, p. 130) and further extended by Turvey and Shaw (1995) is a physics for complex systems that is not relegated to a fundamental level of analysis as mechanistic models are (Machamer, Darden & Craver 2000; Craver & Bechtel 2006). This ecological scale entails a circular causality because the organism and environment alter each other simultaneously, making systems at this scale impredicative (Turvey, 2004; Chemero & Turvey, 2007).

There are compelling reasons to consider the game of football to be a phenomenon of a physical nature, namely that the activity therein is essentially rate (Pattee 1977) and energy dependent (Kondepudi 2008; Kondepudi & Prigogine 2012). Indeed, the very task of football players is to locomote at the right pace towards the right direction and apply forces of the right magnitude and direction relative to the position of ball. The "right" qualifier here adds a normative component which implies knowing the rules of the game, as well as the sorts of strategic actions afforded by the state and evolution of the surroundings. Hence, the task of football is fundamentally one of coupling information to energy, thus requiring something like a physical psychology, viz ecological physics (Turvey & Shaw 1995; López-Felip, 2019). Contrast this to the game of chess. The essence of chess is entirely rate and energy independent. The rules and transitions between states are symbolic and can be realised in multiple different ways. All that needs to be respected is the abstract relationships between the tokens and positions on the board, but nothing in particular is required from a thermodynamic point of view (Haugeland 1985; Brooks 1990).

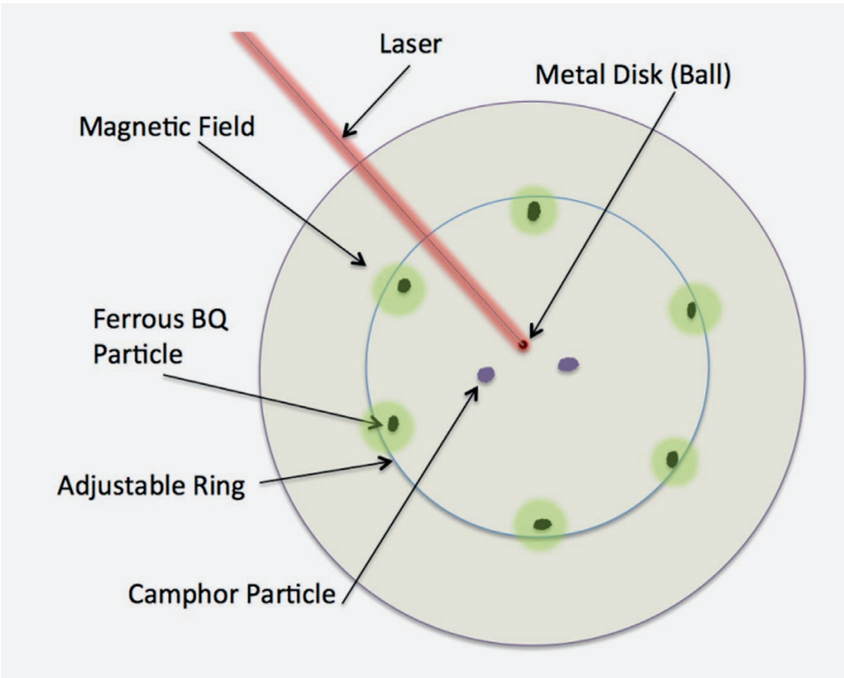
The foregoing considerations entail a critical appraisal of the value of several modeling approaches to the study of football (e.g. Dynamical Systems Theory, statistical physics, etc.). At least two issues can be raised. First, these alone are not sufficient to address the question of how to couple information to energy. Second, even if the tools of these approaches can be quite powerful for modeling, one still needs to make critical decisions about what aspects of a given system one needs to model and how.

Regarding these introductory considerations, this paper aims to provide a glimpse of a research account that has been actively conducting research for more than three years tackling the study of football from that perspective. Specifically, we will explore the relationships between two model systems. One of these is the traditional football exercise of the *rondo*, or man-in-the-middle, which provides a more constrained and self-contained action that nevertheless reveals much about the game. The other model system is one that at a first look, may seem far removed from sports or even human behaviour, but in actuality provides insights into self-organised collective action following the same underlying principles.

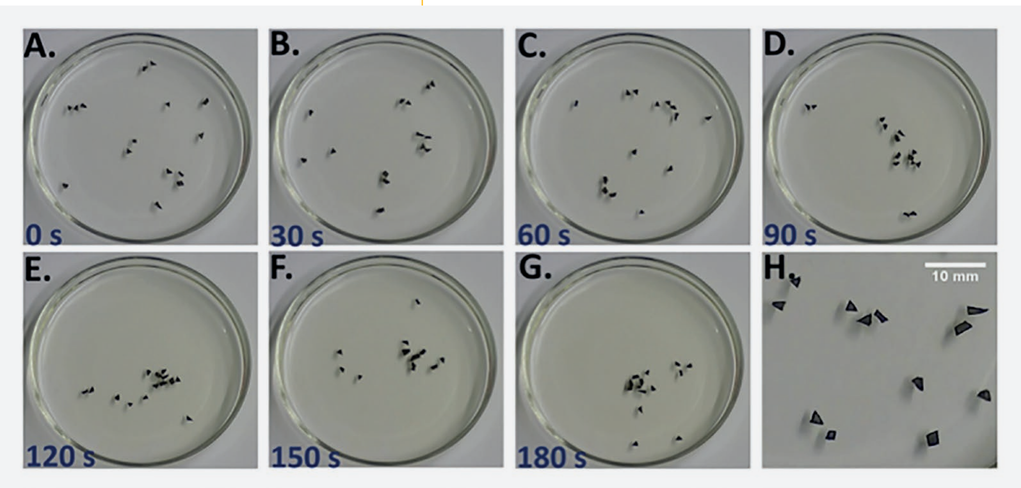
THE C-SOFI PARADIGM

One implication of ecological psychology's emphasis on a physically grounded law-based approach to perceiving and acting is that it is not limited to biological systems (Turvey & Carello, 2012). Indeed, a series of recent findings argue for a fundamental principle for the self-organisation of activity that is indifferent to scale; and that the behaviour manifest in both living and non-living collectives serves to satisfy the dissipation of energy and strives toward the maximisation of entropy production (e.g. Satterwhite-Warden, Kondepudi, Dixon, & Rusling, 2015).

To this end, CESPA has pioneered a model system that can be used as a blueprint to study football from an ecological physics approach. In this system, the organic compound Benzoquinone (BQ) dissolves on the surface of a sodium chloride solution in a petri dish allowing for a significant variety of motion behaviours to be observed. The components of this system, nicknamed the Chemical Self-Organized Foraging Implementation (C-SOFI), are shown in Figure 1. For more methodological details, see e.g. Satterwhite-Warden et al. (2015).



<
Figure 1.
Elements of the C-SOFI system. C-SOFI consists of BQ particles floating at the air-water interface. Even individually, BQ particles exhibit intricate motion due to their interaction with gradients in the medium. Together, they exhibit collective behaviours. In this case, a metal disk heated by a light serves as a catalyst, increasing the energy dissipation of nearby particles.



^
Figure 2. Photographs of a trial of C-SOFI, reproduced from Satterwhite-Warden et al. (2015). Here, we observe flocking behaviour. This self-organised structure serves to dissipate more energy than the particles otherwise would acting independently.

C-SOFI performs spontaneous self-motion due to the interfacial forces generated by their own dissolution into the fluid as a result of changes in temperature. The change of phase that takes place during BQ dissolution has several important consequences, in particular the motion of the particle itself. More importantly, because of radial diffusion in the fluid, when other BQ particles are present the patterns of dissolution rates result in a complex interfacial force field which can lead to emergent collective patterns of attraction and/or repulsion such as flocking, swarming, and so on (Satterwhite-Warden et al. (2015). Previous experiments with this system involved regularly shaped BQ disks (Kondepudi, Kay & Dixon, 2015); irregularly shaped BQ disks (Satterwhite-Warden, Kondepudi, Dixon & Rusling, 2019); temperature changes of the catalyst from hot to cold and vice versa (Satterwhite-Warden et al. 2015); constraints in the environmental layout like a gate (Tianqi, Kondepudi, Dixon, & Rusling F., 2019); combination of regularly and irregularly shaped BQ with a plastic bead (acting as a ball -not published). Results suggest the dissipation of energy and maximum rate of entropy production (MREP) when higher-order (more functional) states of the system are exhibited. An example observation is shown in Figure 2.

Combining principles from dissipative structure theory with the new ontology of the game proposed by FCB's philosophy, we found important analogies between football and the model system. Firstly, analogous to BQ particles, the human athletes that are the components of the football collective propel themselves against the force fields in which they are embedded such as gravity. This shared aspect of the **energetic autonomy** of the components of both systems (the model system and the football system) is what makes the analogy possible.

BQ particles are autonomous given the self-propelling capacity. However, the crucial difference is that BQ particles, although capable of motion through degradation of their own intrinsic free energy, are still driven by interactions with the surrounding force fields directly. Humans, instead, are sensitive to a second-order (informational) field, which implies lower energy and is thus incapable on its own of moving the entire mass of a human being (Shaw & Kinsella-Shaw, 2007). The coupling between information and the internal free-energy is responsible for motion, whereas the BQ system results from energetic interactions of commensurate magnitudes. Nevertheless, the fact that the gradients are informational in one case does not preclude dynamical principles from being operating in analogous ways.

Secondly, considering the **ball** as the element that changes the local energetics of the system through which players move, places the ball in a role analogous to that of a **catalyser**, that is, the hot probe in C-SOFI. Therefore, the assumptions here are that manipulations on the control variable (i.e. ball location) should result in different modes of potential gradient dissipation for players to act so as to dissipate a potential gradient (information) catalysed by the ball.

Thirdly, **end-directness** in the sense that these non-equilibrium dissipative structures seem to prefer states that have higher rates of entropy production and will change their morphology and behaviour to obtain those states (Nicolis & Prigogine, 1977; Kondepudi & Prigogine, 2012; Kondepudi, Kay & Dixon, 2015, 2017). BQ particles in motion increase the rate of entropy production (Satterwhite-Warden et al. (2015), not only generating self-motion, but also producing self-organised motion seen in the flocking behaviour of irregularly shaped fragments (Satterwhite-Warden et al. (2015). In scenarios as complex as team sports, where unique situations are encountered with players sharing intentionality and being, at the same time, highly skilled in the performing activity, one might expect to observe maximum rate of entropy production (MREP) as it can be shown by general considerations (Endres, 2017; Swenson & Turvey, 1991; Swenson, 1997).

Table 1.
The table shows the traits of NESOS that are analogous in both systems. Ball position creates an informational gradient which drives the players according to its potential, as the temperature changes and its potential gradient drives the BQ particles in C-SOFI.
v

C-SOFI	FEATURES	
Temperature change	**Catalyst**	Ball position
BQ particles	**Energy**	Football players
Interfacial field gradient and resulting potential	**End-Directness**	Informational field gradient and resulting potential

THE RONDO

The rondo drill, also known as man-in-the-middle, consists of multiple players forming a circle around some smaller set of interior players. The players on the circle pass the ball amongst themselves, trying to keep it away from the players in the interior. The rondo provides a useful model system in which relevant variables can be more easily manipulated and their effects observed than in the context of a full 11-on-11 match, but nevertheless contains emergent collective behaviour based on ball position.

Specifically, we have previously observed two different collective behaviours with the 2-person rondo (López-Felip, 2019). Depending on the location of the two interior players relative to the ball, they may move closer to the ball such that they equally cover the space in front of the player. We call this pattern side-by-side coverage. With other initial conditions, one of the two interior players closes on the (exterior) player with the ball, while the other interior player stays back to close down the passing lanes (which we call staggered coverage). Neither of these collective behaviours is the result of explicit instruction, but they are clearly quite robust and functional (i.e. they increase the probability of intercepting the ball). Figure 3 shows examples of side-by-side coverage in a small radius rondo, and staggered coverage in a large radius rondo.

^
Figure 3.
An example of each of the behavioural modes exhibited by interior players during manipulations of the radius circle of the rondo in López-Felip (2019). The top picture is an example of a side-by-side mode; the bottom picture is an example of dual-role mode (staggered coverage).

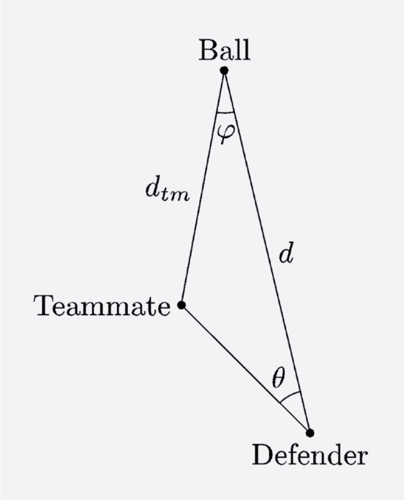
Over the past few years, experimental testing has included the manipulation of variables such as space, number of touches allowed, and number of players, while exploring the resulting patterns using geospatial tracking as well as measurement of energy consumption using VO_2 (López-Felip, 2019). Here, we show some preliminary results from recent modeling efforts focused on the dynamics of two interior players. The ultimate practical application of this line of research is to predict performance exhibited in the rondo and construct a framework for assessing players' collective intelligence.

DYNAMICAL MODEL OF THE RONDO

Our preliminary modeling efforts describe a rondo with six exterior and two interior players. We focus on the interior players, where collective behaviours have previously been observed, modeling their movement as a product of the conditions of the rondo. Currently we treat each possession separately, defined as those brief seconds when one exterior player in particular has the chance to act on the ball. The variables used capture the relationship among the interior players and the ball, as shown in figure 4.

In our model, each player's trajectory, and so the trajectory of the defenders understood as a collective, is generated by the sum of three relationships: (a) the distance between the defender and the ball, mediated by the location of the second defender; (b) the relationship between the two defenders, mediated by the location of the ball; and (c) the passing lanes connecting the possessor of the ball with each of his teammates.

Each of these factors is modeled as a potential, with the defenders influenced by this potential to move in the direction of the negative gradient of the potential.



<
Figure 3.
The variables used in the first two steps of the model.

Defender-ball distance. The ball is modelled as a catalyst that draws the defenders toward it, and its concentration in the field of play is modeled as decaying exponentially. This gives rise to the provisional potential due to the ball:

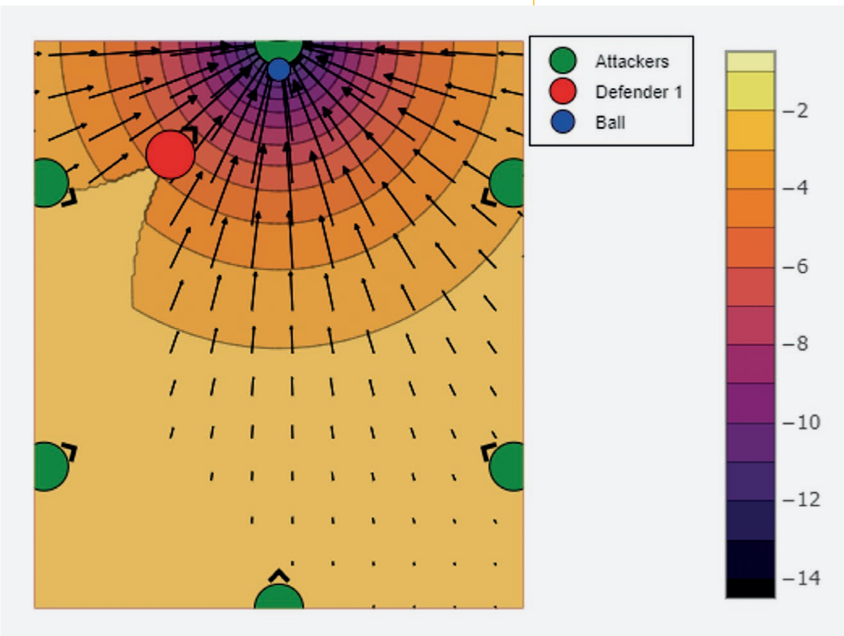
$$V_{ball}=-k_{b1}e^{-k_{b2}d}$$

where k_{b1} and k_{b2} are constants that determine the decay of the concentration of the catalyst, and d is the distance between the defender and the ball. The effect of this potential is that defenders run at a constant speed toward the ball. Because players are modelled as running at a constant speed, if this potential were the only influence on their motion, they would run just as quickly when close to the ball as when far. The lower concentration of the catalyst when the player is further away, however, means that the other factors are more important in determining their trajectory when far from the ball.

In analogy to the physical (BQ) model, in which the catalyst diffuses through an ambient medium, in this model the physical presence of a defender blocks the influence of the ball from diffusing into the area behind it. For this reason, when players are behind their teammate – or more accurately, when a teammate is in the middle 30 degrees of a defender's visual field, they are not directly attracted to the ball and other forces dominate. This "ball shadow" effect, shown in figure 5, leads to the potential due to the location of the ball being expressed by the following equation:

$$V_{ball}=\begin{cases} -k_{b1}e^{-k_{b2}d} & \theta>15^{\circ} \\ 0 & \theta\leq15^{\circ} \end{cases}$$

Figure 5. V_{ball}, and the movement due to it. In this and following figures, darker shades indicate lower potentials. Interior players as modelling as following the gradient toward these lower potentials.
v



Defenders' relationship. Each defender's motion is determined in part by their relation to their teammate. This influence is expressed as a potential itself composed of three components: V_ρ, a function of the difference in the two defenders' distances from the ball, V_θ, a function of the position of the other defender in a defender's visual field and V_φ, which is a function of the angle formed by line segments connecting each player's location with the location of the ball.

V_ρ is defined so the defenders act as if a spring pulls on them to be the same distance from the ball as their teammate. The implied spring constant is a function of how central the player's teammate is in their visual field, and whether they are closer to the ball than their teammate (and so, as it were, unable to see them). The potential, therefore, is modeled as:

$$V_\rho = \begin{cases} k_{\rho 1}(d-d_{tm})^2 & \theta > 15° \text{ and } (d-d_{tm}) > -0.5 \\ k_{\rho 2}(d-d_{tm})^2 & \theta \geq 15° \text{ and } (d-d_{tm}) > -0.5 \\ 0 & (d-d_{tm}) < -0.5 \end{cases}$$

In simulations, $k_{\rho 1}$ was set to 1, and $k_{\rho 2}$ was set to 0.01. The resulting potential is shown in the left panel of figure 6.

V_θ is defined, by analogy to a spring that pulls the angle made by the segments connecting the player to his teammate and the player to the ball toward 90 degrees, although with constant force so it still pulls when theta is close to 90 degrees. V_θ therefore, is given by the following equation (illustrated in the right panel of figure 6):

$$V_\theta = \begin{cases} k_\theta \left|\theta - \frac{\pi}{2}\right| & \text{when } \theta > -60° \\ 0 & \text{otherwise} \end{cases}$$

Finally, the potential due to φ is defined by analogy to a spring that pulls the angle made by the segments connecting each player to the ball toward 30 degrees ($\pi/6$).

$$V_\varphi = k_\varphi \left(\theta - \frac{\pi}{6}\right)^2$$

The sum of these three potentials can be expressed as the total direct influence of the teammate's position on a player's movement:

$$V_{tm} = V_\rho + V_\theta + V_\varphi$$

This combined potential is graphed in figure 7.

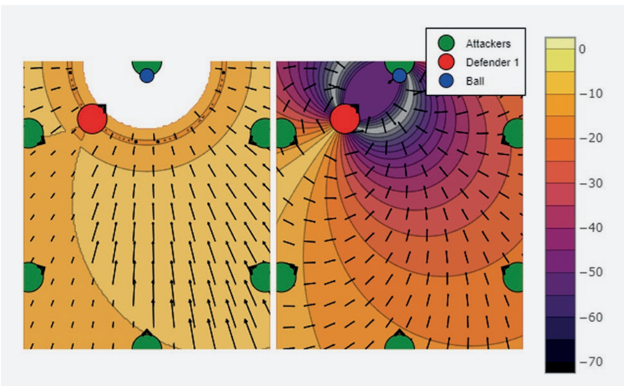


Figure 6. V_ρ (left) and V_θ (right).

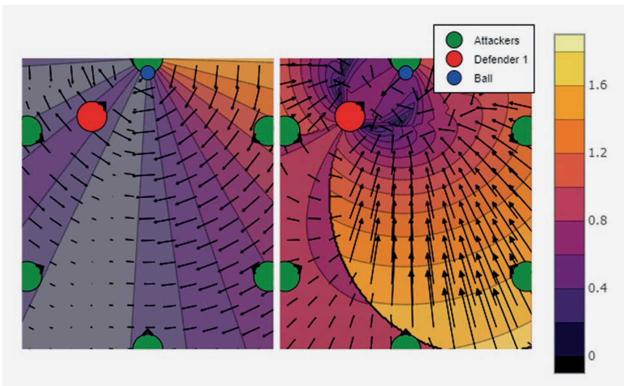
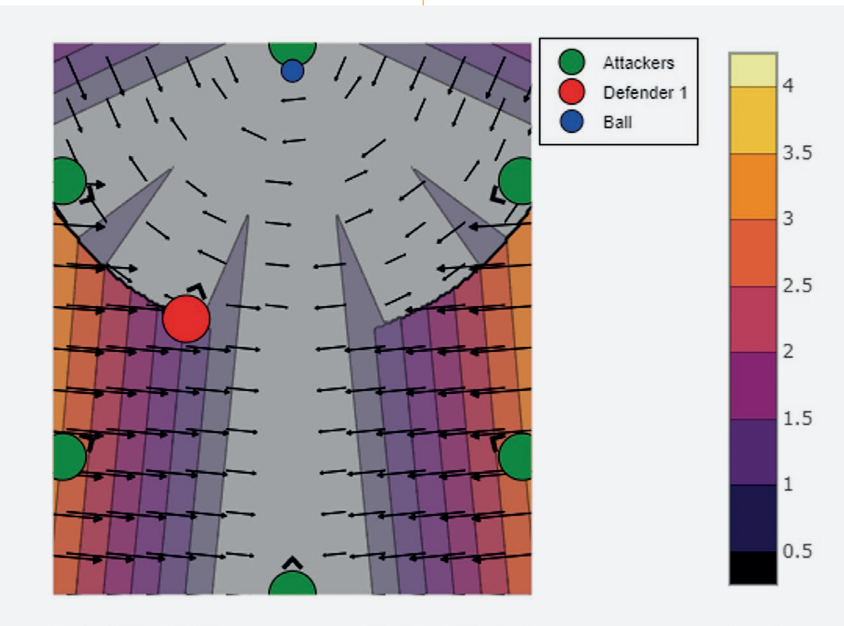


Figure 7. V_φ (left), and $V_{tm}=V_\rho+V_\theta+V_\varphi$ (right).



^
Figure 8: V_{lane}. If defender 2 is in front of defender 1, they move toward the nearest passing lane. If they are behind defender 1, they move to cover the passing lane to defender 1's right, since it is adjacent to defender 1, and on the open side.

Passing lanes. Defenders attempt to coordinate their activity to cover passing lanes. Specifically, the defender who is closest to the ball moves as if they are attached by a spring to the nearest passing lane, and the defender who is furthest from the ball identifies the passing lane their teammate is closest to covering, and moves as if attached by a spring to the adjacent passing lane, on the more open side. If their teammate is covering the lane to the attacker directly opposite the ball, the defender covers the nearest adjacent passing lane. Because players have more time to intercept a pass when they are further away, the width of a passing lane increases linearly with distance from the ball. This results in the following equation:

$$V_{lane}=\max[k_{lane1}\,(d_{lane}-k_{lane2}\,d)^2,0]$$

where d_{lane} is the distance to the appropriate lane, d is the distance from the defender to the ball, k_{lane1} is the spring constant, and k_{lane2} expresses the rate the width of the passing lane increases. Figure 8 illustrates which lane a defender is attracted to.

The total potential, V, is then the sum of the three components, and the direction of motion is determined by the negative gradient of the potential.

$$V= V_{ball}+V_{tm}+V_{lane}$$

Motion is in the direction $-\nabla V$.


14
STANLEY
unicef

RESULTS

Previous studies (López-Felip, 2019) identified three different modes of defender behaviour: (a) both defenders pressing the ball, (b) a dual-role mode in which one defender presses the ball and the other covers passes, and (c) a mode in which both defenders cover passes. This model was able to qualitatively reproduce all three modes. These three modes are illustrated in Figure 6a–c. In all of these images, time was run for a quarter of the time it takes an agent to run in a straight line from one attacker to the attacker opposite the first, to illustrate the emergent tactical coordination of the defenders. In Figure 9a, rather than running directly toward the ball, the defenders have coordinated their positions to cover two passing lanes. In Figure 9b, the forward defender began by pressing the ball, while his teammate moved behind him to cover the passes. Finally, in Figure 9c, the defenders coordinate their position, and press the ball side-by-side.

This behaviour can also be compared to data collected from rehearsals by players from FCB team 2 (as analysed in REF). Figure 10 shows stills from a video of FCB team 2 members participating in rondo drills as part of a prior study. In the upper panel, the ball started on the right, and the defenders adopted a side-by-side cover strategy, as seen from the screen-shot just after the pass. At that point, the defenders again adopted a side-by-side cover strategy, and then on the two subsequent passes, they adopted a dual-role mode, with one defender pressing the ball and their teammate covering a pass.

Figure 9. A) Left: Side-by-side cover. **B)** Centre: Dual mode, one player presses, and the other covers. **C)** Right: Both players press the ball.
v

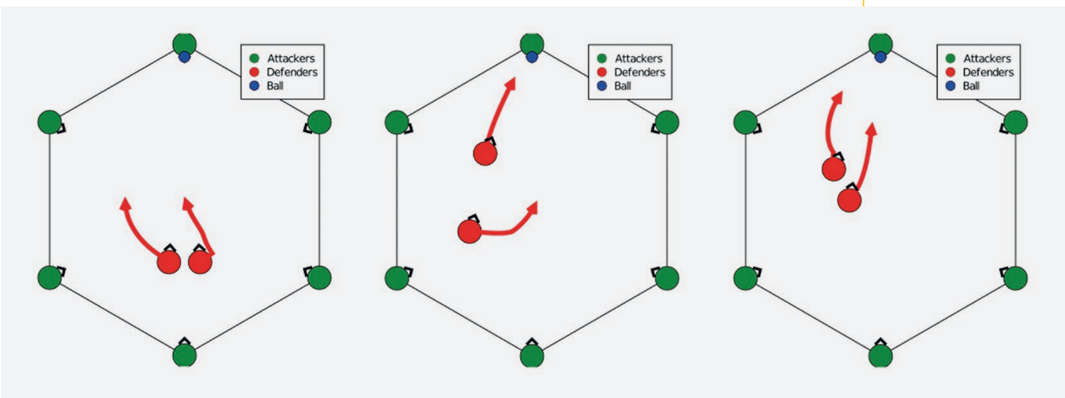
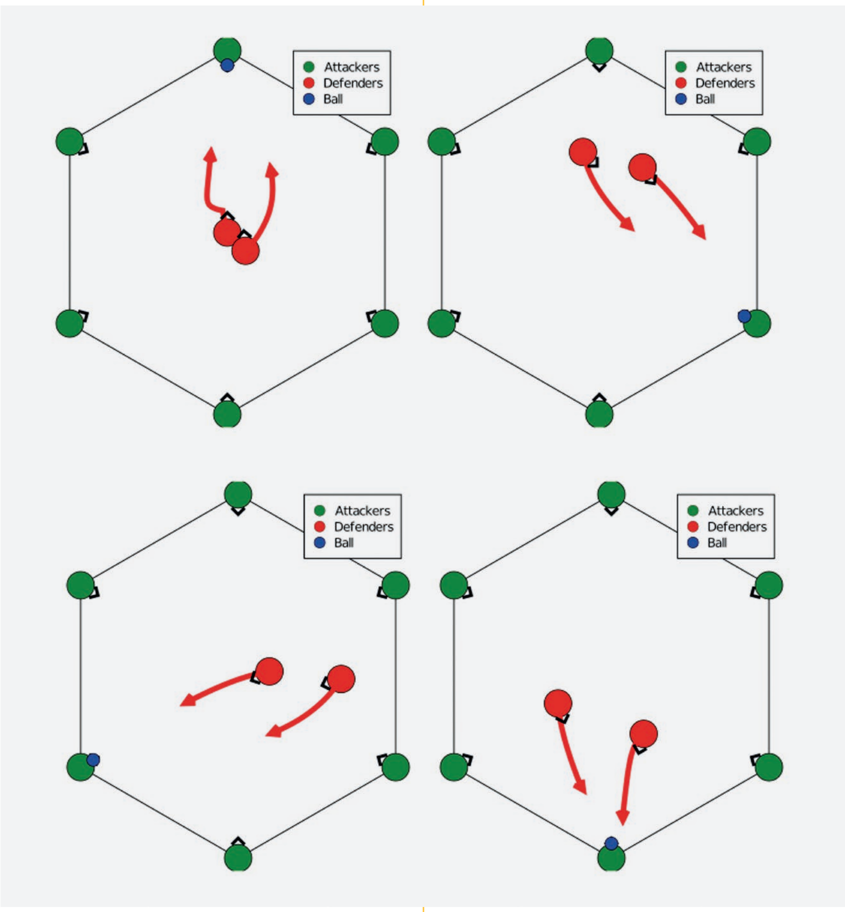


Figure 10. A simulation of the same sequence of passes. Except for the last image, modeled behavior is qualitatively similar to the data collected from members of the FC Barcelona team 2 drill.
v



In a simulation (Figure 11) defending unit followed a similar (though not identical) trajectory: At first, they adopted a side-by-side covering strategy (this time toward the top of the hexagon). At first, they adopted a side-by-side covering strategy (this time toward the top of the hexagon). After the pass they seem to have adopted a side-by-side strategy, though here the model could be interpreted as dual-mode strategy. Following a second pass, they adopted a dual mode strategy, with the player on the left (in the diagram) running toward the ball while the player on the right covered the open passing lane. Finally, in the final figure, both defenders pressed the ball.

Figure 10. Four stills from a study of individuals' participation in a rondo.

DISCUSSION

We have shown that we can reproduce some patterns observed in the rondo by composing simple behavioural rules focused on the relationship of the interior players to the ball. Basic considerations of the rondo allow us to construct hypothetical informational gradients and roughly approximate their potentials in order to observe the behavioural patterns that emerge from these potential fields.

With respect to the modelling effort, two major issues remain to be addressed: First, although we can recreate real sequences by simulating successive initial conditions, we have only modelled defensive behaviour during one attacker's time in possession. For this reason, the model treats an attacker as possessing the ball indefinitely, and, as if an attacker were unwilling to pass, if time is allowed to go to infinity, the defenders eventually begin to press the ball. Future models need to address passes more explicitly.

One line of approach can be seen from observing that this problem could equivalently be phrased as: According to the model, passes are made instantaneously, and defenders are incapable of anticipating a pass till it is made. But once initiated, the act of passing the ball takes a relatively predictable amount of time, and even before a pass has been initiated the time to pass is relatively predictable. Therefore, any model

parameters should be a function of time-to-kick. Our goal therefore for future research, is to incorporate a specific time-to-kick variable in the model, and to allow the parameters in this model to vary with time. Modeling the rondo over multiple sequences will also require a model of pass selection on the part of the outer players. Here a physics-based model (e.g. Spearman, Pop, Basye, Hotovy & Dick, 2017) can be incorporated.

Second, the parameters in this model have not yet been tuned to data—indeed, if as the last paragraph conjectured, they are a function of time-to-kick, they cannot be tuned till time-to-kick is incorporated into the model. The final step of this work, therefore, will be to tune the parameters as a function of time-to-kick.

Besides these modeling considerations, future work should continue to investigate the analogy between collective behaviour in football and the observations of the C-SOFI experiment. Here, we connect the two systems in terms of the modeling approach and assessment of the components of each system. In the future, the analogy can be made more concrete by replicating the empirical and statistical findings from C-SOFI in the context of the football rondo.

Specifically, previous work in C-SOFI indicates that functional collective behaviour such as flocking increases the rate of entropy production. This observation is consistent with dissipative structure theory, as the collective structure facilitates the dissipation of energy. Therefore, we could expect similar patterns of results in the rondo. Once this is confirmed, the level of coordination of any pair of players could be assessed in terms of entropy production. We would hypothesise that those players exhibiting greater entropy production would be forming more functional collective structures, an outcome which could have implications for assessing performance in actual matches.

Indeed, developing an assessment methodology based on the principles of self-organising systems would provide unique insight into real-world performance. Whereas other novel assessment methods may improve player and team evaluation in terms that coaches are already familiar with, the one proposed here would provide a new perspective and therefore be particularly complementary if added to the suite of tools at the disposal of a modern coaching staff.

More generally, the overarching goal of the research project is to better understand the emergence of collective behaviour, grounded in dissipative structure theory. Biological agents routinely configure themselves into special-purpose physical devices in the service of their immediate goals. Despite their ephemeral nature, these special-purpose physical devices have real physical consequences and implications (e.g. they generate forces, consume resources, and alter local fields). When many such physical devices interact in a field, the system will quickly find the configuration (i.e. set of dynamics) that satisfies the physical laws. These considerations have revolutionised our understanding of human behaviour but have yet to make an impact in the study and practice of football. Having principled scientific evidence of how functional modes appear in the game and being able to quantify which ones are more efficient regarding the actual context of the game will allow us to pursue a talent development methodology that is driven by physical principles in combination with coaching insights tailor-made by FCB.

REFERENCES

Brooks, R.A., Intelligence without representation, *Artificial Intelligence*, 47 (1991), 139–159.

Chemero, A., & Turvey, M. T. (2007). Complexity, hypersets, and the ecological perspective on perception-action. *Biological Theory*, 2, 23–36.

Craver, C. & Bechtel, W. (2006). Mechanism. (S. Sarkar & J. Pfeifer, Eds.). *Philosophy of Science: An Encyclopedia*. New York: Routledge.

Davis, T. J., Kondepudi, D., Kay, B. A., Dixon, J. A. (2016). Spontaneous inter-entity coordination in a dissipative structure. *Ecological Psychology,* 28, 23–36.

Dixon, J. A., Kondepudi, D., Kay, B. A. & Davis, T. J. (2016). End-directedness and context in non-living dissipative systems. In E. N. Dzhafarov (Ed.), *Contextuality from Quantum Physics to Psychology* (pp. 185-206). World Scientific Press, Hackensack, NJ.

Endres R. G. (2017). Entropy production selects nonequilibrium states in multistable systems. *Scientific reports*, 7, 14437. https://doi.org/10.1038/s41598-017-14485-8

Frank, T.D. (2015). On the Interplay between Order Parameter Dynamics and System Parameter Dynamics in Human Perceptual-Cognitive-Behavioral Systems. *Nonlinear Dynamics Psychol Life Sci*. 19:111-46. PMID: 25783041.

Gibson, J.J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge, MA: MIT Press

Kondepudi, D. (2008). *Introduction to modern thermodynamics*. New York, NY: Wiley.

Kondepudi, D., & Prigogine, I. (2012). *Modern thermodynamics, from heat engine to dissipative structures*, 2nd ed. New York, NY: John Wiley.

Kondepudi, D., Kay, B., & Dixon, J. A. (2015). End-directed evolution and the emergence of energy-seeking behavior in a complex system. *Physical Review E*, 91, 050902.

Kondepudi, D., Kay, B., & Dixon, J. (2017). Dissipative structures, machines, and organisms: A perspective. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27, 104607.

López-Felip, M. A., Davis, T. J., Frank, T. D., & Dixon, J. A. (2018). A cluster phase analysis for collective behavior in team sports. *Human movement science*, 59, 96-111.

López-Felip, M.A. "Collective Behavior in Dissipative Systems: Flocking & Futbol" (2019). Doctoral Dissertation. 2316.

Machamer, P., Darden, L. & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*. 67, 1–25.

Nicolis, G., & Prigogine, I. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.

Pagano, C., & Turvey, M. T. (1998). Eigenvectors of the inertia tensor and perceiving the orientations of limbs and objects. *Journal of Applied Biomechanics*, 14, 331-359.

Pattee, H. (1977). Dynamic and linguistic modes of complex systems. *International Journal of General Systems*, 3, 259-266.

Rosen, R. (1991). *Life itself*. New York: Columbia University Press.

Satterwhite-Warden, J.E., Kondepudi, D.K., Dixon, J.A., & Rusling, J.F. (2015). Cooperative motion of multiple benzoquinone disks at the air–water interface. *Physical Chemistry Chemical Physics*, 17, 29891-29898.

Satterwhite-Warden, J. E., Kondepudi, D. K., Dixon, J. A., Rusling, J. F. (2019) Thermal- and Magnetic-Sensitive Particle Flocking Motion at the Air-Water Interface. *Journal of Physical Chemistry B*, 123, 3820-3840.

Shaw, R. & Kinsella – Shaw, J. (1988). Ecological Mechanics: A Physical Geometry for Intentional Constraints. *Human Movement Science*, 7 155 – 200.

Shaw, R. E. & Kinsella – Shaw, J. (2007). Could Optical "Pushes" Be Inertial Forces? A Geometro-Dynamical Hypothesis. *Ecological Psychology,* 19, 305 – 320.

Spearman, W., Pop, P., Basye, A., Hotovy, R. & Dick, G. (2017) Physics-based modeling of pass probabilities in soccer. In: *Proceedings of the 11th MIT Sloan Sports Analytics Conference*, pp 1–14.

Swenson, R. & Turvey, M. (1991). Thermodynamic Reasons for Perception-Action Cycles. *Ecological Psychology*, 3, 317-348.

Swenson, R. (1997). Autocatakinetics and the law of maximum entropy production: A principled foundation towards the study of human ecology. *Advances in Human Ecology*, 6, 1-47.

Tianqi, C., Kondepudi, D. Dixon, J. & Rusling F. (2019). Particle Flock Motion at Air–Water Interface Driven by Interfacial Free Energy Foraging. *Langmuir*, 35, 11066–11070

Turvey, M. T. (2004). Space (and its perception): The first and final frontier. *Ecological Psychology*, 16, 25-29.

Turvey, M. T., & Carello, C. (2012). On intelligence from first principles: guidelines for inquiry into the hypothesis of physical intelligence (PI), *Ecological Psychology*, 24, 3-32.

Turvey, M. T. & Fonseca, S. T. (2014). The medium of haptic perception: A tensegrity hypothesis. *Journal of Motor Behavior*, 46, 143-187.

Turvey, M. T., & Shaw., R. E. (1995). Toward an ecological physics and a physical psychology. In R. L. Solso, & D. W. Massaro (Eds.), *The science of the mind: 2001 and beyond*. New York: Oxford University Press.

Beyond 'blind' distance covered in football match analysis: is it time to progress to a contextualised paradigm?

— Paul Bradley[1], Wonwoo Ju[2], Jack D. Ade[3], Andy Laws[2], Antonio Gómez-Díaz[4] & Mark Evan[5]

1: Football Science Consultant

2: Researcher in Football Science at LJMU

3: U23's Fitness Coach at Liverpool Football Club

4: FC Barcelona

5: Computer and Data Scientist

INTRODUCTION

A piece was recently published in the *International Journal of Sports Physiology and Performance* that has potentially heralded a new contextualised way of analysing, interpreting and most importantly translating football match running performance data to the coach and practitioner.

Since the seminal work of Reilly and Thomas (1976), the match running performances of elite football players have been typically analysed as accumulated distances, frequencies and times in various arbitrary motion categories. Across these 40+ years, hundreds of papers have been published using such an approach, including many works by the authors of this 2020 FCB Football Analytics piece (Bradley et al. 2009) Barnes et al. 2014; Da Mota et al. 2016; Martín-García et al. 2018a). Many interesting and novel findings have sprung from this archive, such as the evolution of the game (Barnes et al. 2014; Bush et al. 2015; Bradley et al. 2016; Bradley & Scott, 2020; Norton and Wallace, 2014), positional variation (Da Salvo et al. 2009), reductions in running performance (Mohr et al. 2003), match-to-match variation (Bush et al. 2014); Carling et al. 2016; Gregson et al. 2010 to name just a few. Although many still passionately discuss the appropriate speed thresholds, dwell times (Varley et al. 2017 and the inclusion of much needed and controversial metrics (Akenhead et al. 2013; Osgnach et al. 2010), are we missing the translation and narrative surrounding such analyses? Is it not the sole aim of expending valuable time and financial resources on this area to provide information to the end users that they can use and understand?

These are complex questions but simplifying this to a few salient points could help to innovate and develop the area in the future and use the valuable commodity of time more effectively and efficiently. The term 'blind' distance covered was used to describe the ground covered by players without considering context (Bradley et al. 2018). It is concerning that the cutting edge of the industry has continued to follow this traditional method without questioning or innovating the 'blind' match analysis culture. Various stakeholders vary in their knowledge of the science and the game itself, so it is understandable why this persists. The root cause of the problem, stems from the reductionist nature in which the physical metrics are explored with little or no consideration for the technical and tactical indices (Akenhead et al. 2013; Bangsbo et al. 1991; Bradley et al. 2009; Di Mascio & Bradley, 2013; Di Salvo et al. 2009). Thus, using such an approach is akin to effectively analysing a multifaceted sport like football using one-dimensional tools. Match running performance data are simply numbers but adding context provides a much needed narrative to the numbers, especially if football language and terms are integrated.

This FCB Football Analytics piece will examine some of the contextualised match analysis work our group have conducted at the cutting edge of the English Premier League. Special emphasis will be placed on how such innovative analyses are translated to the end user.

SIMPLE PARADIGM OF FOOTBALL PERFORMANCE

Before providing an alternative approach when analysing football performance, it is imperative that some basic framework is established to construct ideas and refine the initial concept. Figure 1 depicts a basic Venn diagram of elements related to football performance. Three performance factors are represented in isolation and combination as circles (Bradley and Ade, 2018). Please note that these factors are not exhaustive, as elements like player psychology are vital to performance but the three included can be quantified more readily. The regions where factors overlap are the intersections. The area whereby all factors overlay is called the union (black dot) and denotes major innovation in match analysis as full integration occurs. The union of these facets could be considered beyond the realms of technology and expertise at present.

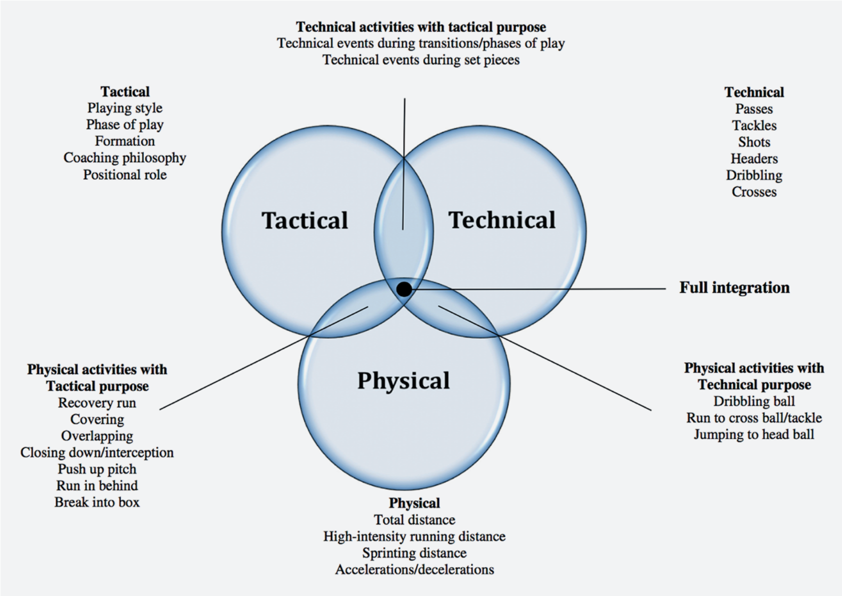


Figure 1. A Venn diagram depicting a generalised 'integrated' approach to quantifying and interpreting the physical match performance of football players. This approach focuses on high-intensity running efforts across the game but contextualises these actions in relation to key technical and tactical activities. Taken from Bradley and Ade (2018).

This piece will focus on the intersection of the Venn between physical-tactical factors. The variables listed within this intersection were created from a 'High Intensity Movement Programme' (Ade et al. 2016) and more recently refined for industry by Ju et al. (unpublished data). The new data sets used in the examples below comprised of teams tracked during elite matches using a computerised system. High-intensity efforts were activities reaching speeds >19.8 km·h-1 for a minimal dwell time of 1 second. To contextualise data, the tactical actions associated with each effort were manually coded from video recordings viewed using bespoke software designed for industry purposes. Please refer to previous research for details of the definitions for the tactical-physical actions and zonal areas used (Ade et al. 2016; Bradley and Ade, 2018). In 2019, work was undertaken to verify the validity of the new approach and to apply this contextualised paradigm to elite clubs (Ju et al. unpublished). In fact, 30 elite performance analysts and UEFA qualified coaches were asked to identify random video clips that were pre-defined by an expert panel as predominantly illustrating a set action. This resulted in ~90% agreement between the expert panel that selected the clips and the coaches and analysts who participated, indicating the validity of the approach. Thus, they see what the experts see!

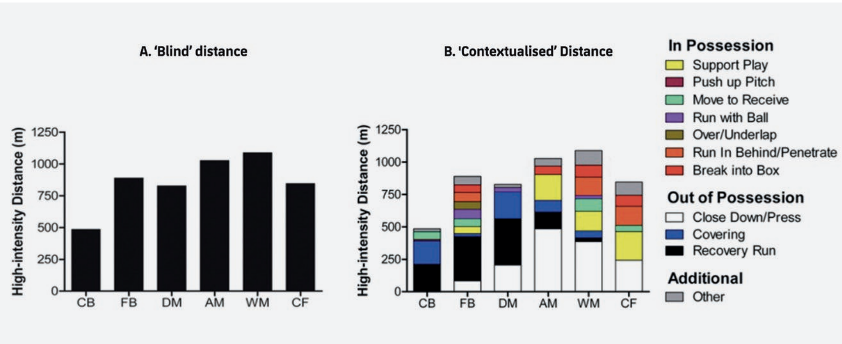


Figure 2. High-intensity distance covered using the **(A)** 'blind' distance covered versus **(B)** 'contextualised' distance covered. The bottom of each contextualised stack includes out of possession variables while the top includes in possession variables for each position. CB=Centre Back; FB=Fullback; DM=Defensive Midfielder; AM=Attacking Midfielder; WM=Wide Midfielder; CF=Centre Forward. These are based on the primary context but some hybrids are included. Taken from Ju et al. (unpublished).

TRANSLATION OF CONTEXTUALISED MATCH RUNNING PERFORMANCE: INSIGHTS

POSITIONAL PHYSICAL-TACTICAL DATA

One of the most consistent findings within the research literature is that the physical demands of training and match play are highly dependent on the players' tactical role (Di Salvo et al. 2009; Martin Garcia et al. 2018b). Coaches and practitioners do not only want to know 'WHAT' distance was covered but 'HOW' and 'WHY' each player covered that distance to better understand the performance of specific duties in relation to the opponent/team philosophy (Bradley and Ade, 2018). Contextualised physical data that merges high-intensity running with the tactical purpose of the action may provide confirmation that players are abiding by tactics or the game plan. Figure 2 compares effectively the 'WHAT' versus the 'WHY' or the 'blind' versus the 'contextualised' distances. This data was taken from a single English Premier League match in which players from the home and away team were randomly selected based on their playing position. As can be seen, the 'blind' approach is rudimentary compared to the 'contextualised' model. The latter illustrates that offensive positions 'press/close down' more than the players with dual or defensive roles. This is understandable as both teams tried to defend from the front by delaying/pressurising high up the field. This is a clear reminder to offensive positions of their duties in both attacking and defending phases.

The defensive duties of 'recovery run' and 'covering' are highest for centre backs, fullbacks, and various midfielders with offensive players conducting less of this activity. 'Support play' is very common when the ball is advanced forward rapidly and the players behind the ball run offensively at high-intensity to provide assistance. As this type of assistance is vital for producing a viable offensive threat this is viewed positively by applied staff. As this game was very condensed at times in central pitch zones, wide players like fullbacks and wide midfielders produced intense efforts while 'running with the ball' in wide zones. Bespoke activities like 'under/overlapping' were conducted primarily by fullbacks in this game. The actions 'run in behind' and 'break into box' are produced more often by the most offensive players in the game. Linkages between variables may also be discerned. As the fullback selected was very offensive as evidenced by 'run in behind' and 'break into box' distance, he was generally higher up the pitch in possession. However, when turnovers in possession occurred during the game, the fullback was required to produce extensive 'recovery runs' to maintain his sides defensive integrity, hence the sizeable composition for this defensive duty.

INDIVIDUAL PHYSICAL-TACTICAL DATA

Visualising this data from a 'birds-eye view' also provides added perspective to coaches who are interested in specific individual performances. Figure 3 visualised a wide forward's contextualised high-intensity running performance during an English Premier League match (Ju et al. unpublished). Imagine the 'blind' equivalent of these colour coded arrows: they would simply be arrows with no narrative or meaning!

Figure 3 illustrates that the wide forward performed 45% of his purposeful actions in the form of 'run in behind/penetrate' and 'close down/press' which are very desirable for his team's playing style (e.g. a high pressing, counter attacking side). 'Closing down/pressing' high up the field in the first half is very evident. As he is the right sided forward of a front three, it is easy to visualise how he has attempted to 'pin' and delay the opposition player(s) into the corner to enable teammates to support the press. The majority of the efforts for these two categories are very short and explosive in nature and typically produce some of the most intense sequential accelerations and decelerations compared to other variables (Ju et al. unpublished). Imagine trying to discern the tactical purpose from a 'blind' approach!

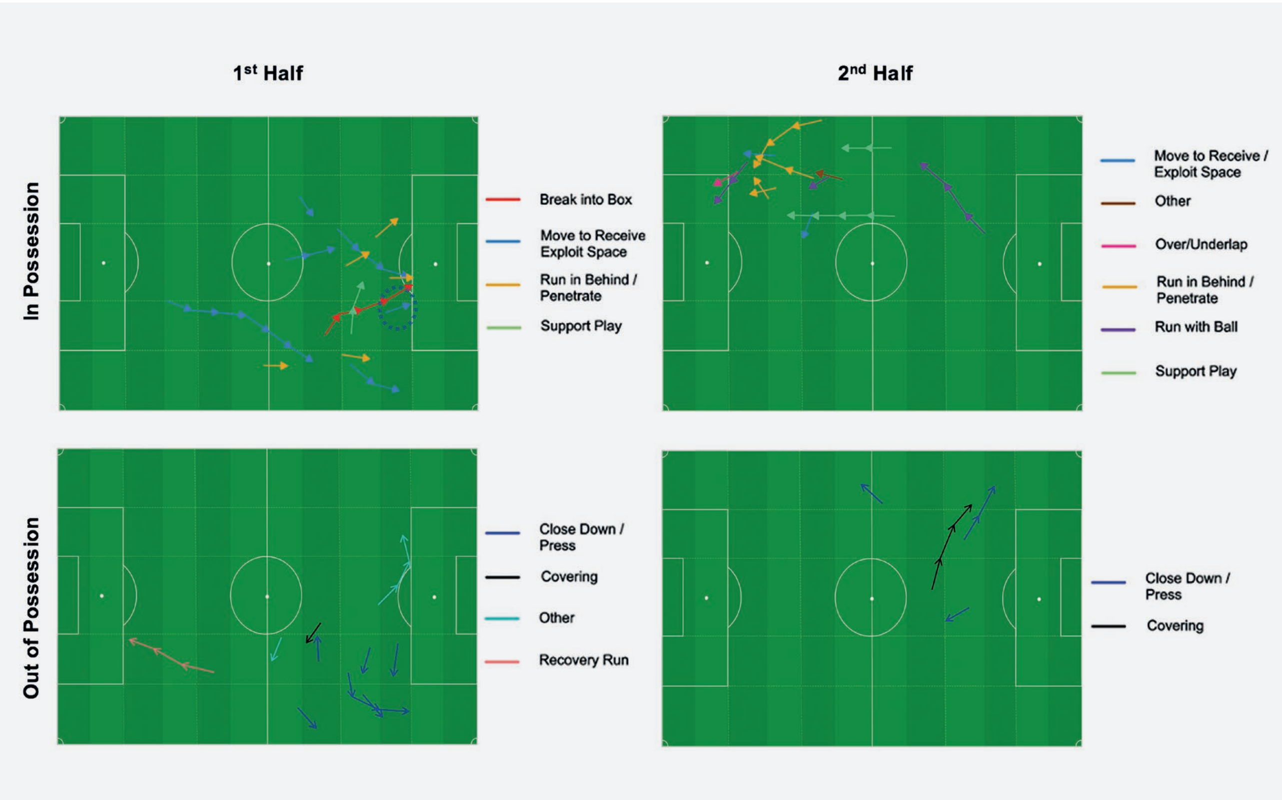


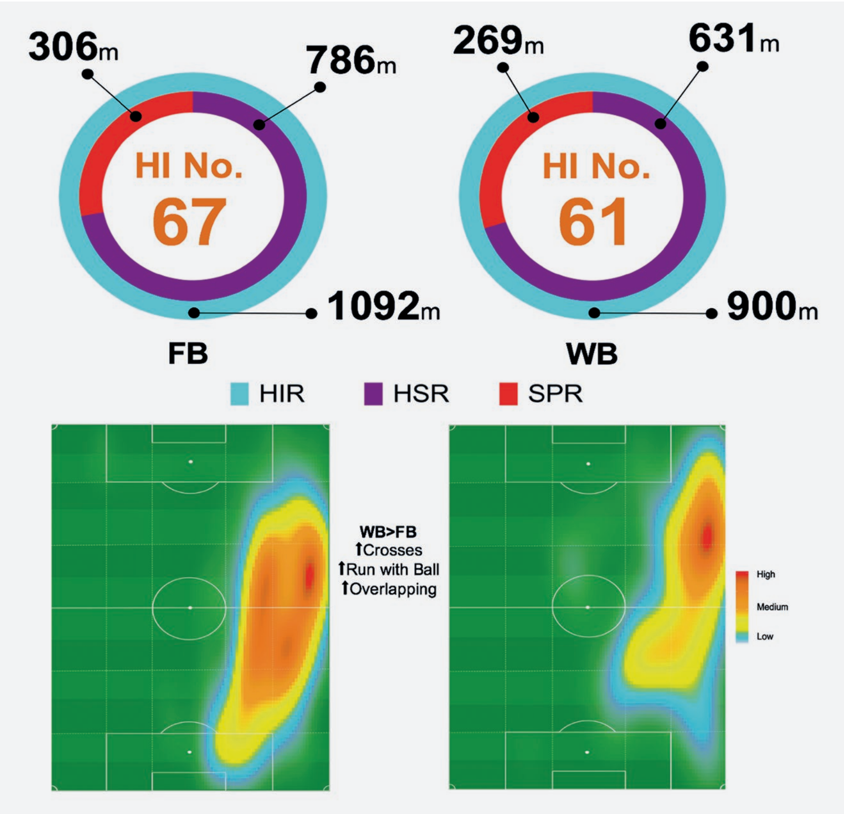
Figure 3. An example of an English Premier League wide forward's high-intensity running actions combined with the colour coded 'primary' tactical context. To help gauge changes in speed each arrow represents a 1 sec period. Raw tracking data was visualised using the "ggplot2" package for the R statistical programming language. These are based on the primary context but some hybrids are included. Taken from Ju et al. (unpublished).

'Move to receive/exploit space' made up 20% of all his efforts and he produced his longest run of the game in this category. This occurred during a counter attack from an opposition corner in the first half where he positioned himself to receive the ball before it was intercepted by the opposition. Interestingly, the 'move to receive/exploit space' action highlighted with a dotted blue circle led to him receiving a pass outside the box before shooting to score the winning goal. The forward only produced a single intense action in the category of 'break into the box' in which he switched the play by passing into an advanced wide area for the wide defender to cross. After the pass, he then accelerated rapidly towards the box but as he approached the area he slowed down in anticipation of the cross. This ultimately led to an attempt on goal for his teammate. Only three bouts of 'running with the ball' at high-intensity were recorded for this player. Two of the three are of interest, as the longest 'run with the ball' was again during another counter attack from an opposition corner. The most advanced 'run with the ball' action occurred when he dribbled into the box at high-intensity before unleashing a shot on goal that went wide. The applied staff would also be delighted about the forward's unselfish behaviour in the first half as he produced a long 'recovery run' to interchange for a teammate out of position. All of this adds a clear narrative to his performance by unveiling the running profile that exists due to his unique tactical role in the team, rather than one-dimensional 'blind' distances that have limited context (Bradley et al. 2018). For practitioners, capturing contextualised data during a period of successive games can provide enough information to design the optimum simulated training situations in order to replicate the most demanding physical and tactical events.

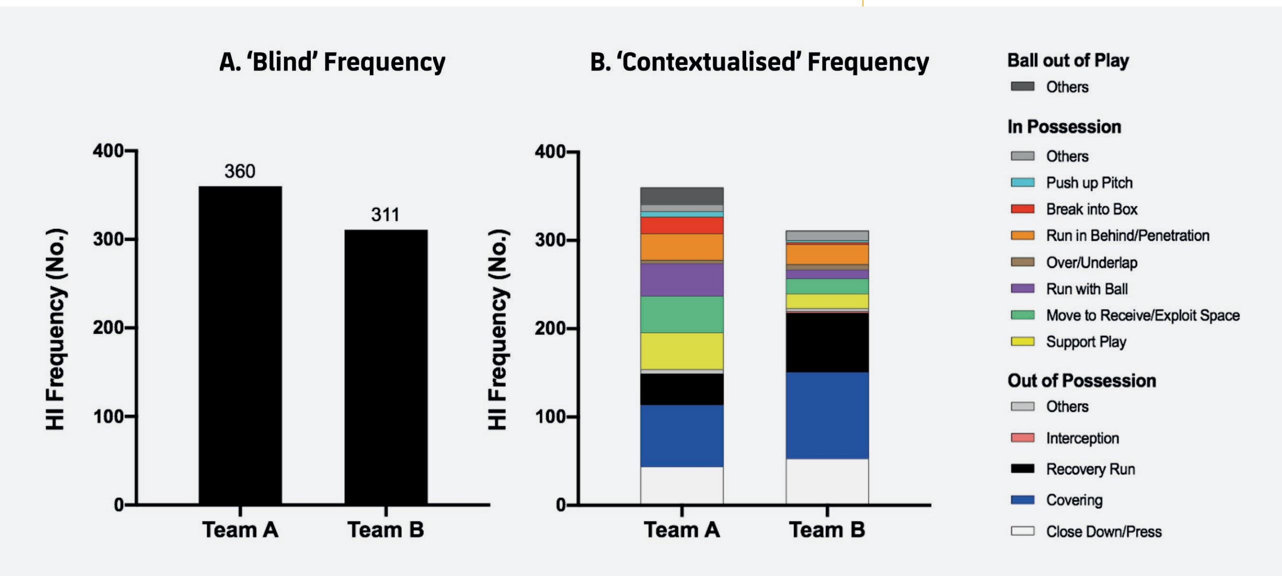
Other applications for such an individualised contextualised approach could be to visualise how a change in the tactical system impacts the physical-tactical profiles of selected players. Specifically, incorporating a back three to deploy the wide defenders as attacking wing backs (WB) or moving to a back four and using them as traditional fullbacks (FB). The example in Figure 4 not only depicts the overall high-intensity running of this scenario or 'blind' distance covered but also breaks this down into the high-speed running and sprinting subsets via colours to delve deeper into the specifics. Although helpful, it is very one-dimensional as it lacks any contextual linkage and thus the visual below depicts his high-intensity running frequency in relation to pitch zones as a heat map with the context of selected actions added. In an instant, this informs the coach of any similarities/differences between external defenders based on tactical shifts (e.g. some difference in absolute distances and number of efforts but particularly the area deployed and the context of technical/tactical actions differs). However, longitudinal data needs to be presented for a more complete picture and the match-to-match variation needs to be added to the narrative, so one can decipher the signal from the noise (Bush et al. 2015; Carling et al. 2016). The authors are not stating that 'blind' distance is not informative (WHAT), far from this but by combining it with context (HOW and WHY) potentially enhances the metrics value.

TEAM PHYSICAL-TACTICAL DATA

As football is a team sport, collective performances are of real importance to the support staff as they can establish if the game plan was adhered to. Thus, this contextualised methodology can provide added insight regarding team physical-tactical performances. Figure 5 illustrates the frequency of intense efforts during an English Premier League match between the home (Team A) and away teams (Team B) using the 'blind' versus 'contextualised' approaches. As can be observed in the 'blind' analysis, Team A conducted ~50 extra intense bouts in the game compared to the Team B (WHAT) but only the additional granular information can be obtained from the 'contextualised' output (WHY). Due to Team A dominating possession and offensive play, the away team carried out ~100% more 'recovery runs', in addition to >20% more 'covering' and 'closing down/pressing'. These actions were conducted to delay, nullify and dispossess Team A to enable a counter attack to occur. In contrast, Team A produced between 30-270% more high-intensity bouts 'running with the ball', 'moving to receive/exploit space', 'running in behind'. The associated collective assistance in the form of 'support play' was ~150% higher. The largest difference was for the high-intensity effort of 'breaking into the box', which occurred 19 times for Team A and only twice for Team B. Although game outcomes are complex (Collet, 2014), it is not surprising that Team A won this game by a goal differential of two and absolutely dominated offensively. Thus, the contextualised approach presented here is useful and valuable for positional, individual and collective insights.



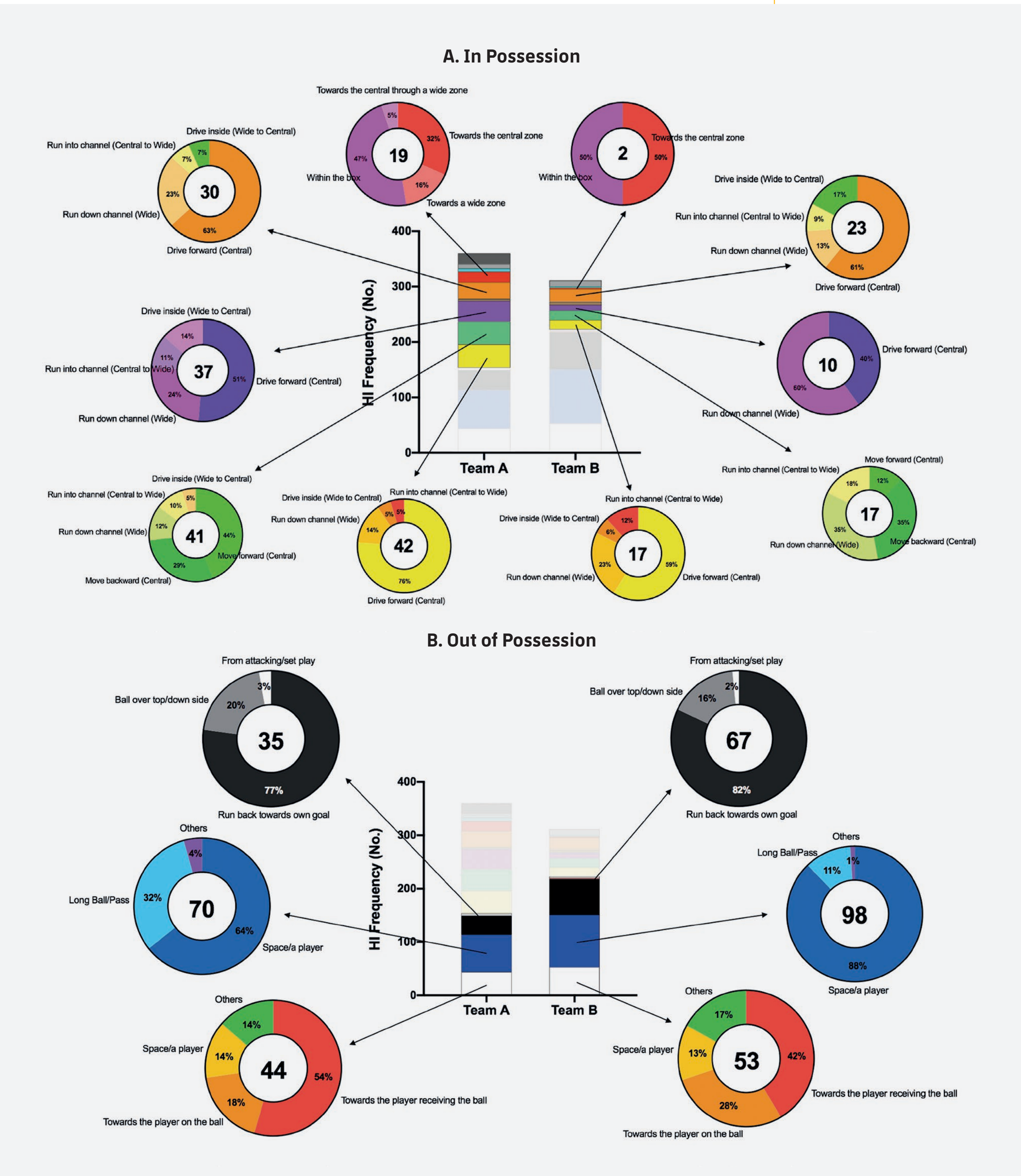
<
Figure 4. The impact of tactical changes in the English Premier League on a wide defender's high-intensity activity. Data are visualised as a stacked donut chart (frequency in the centre and distances as circles) combined with heat maps (frequencies per pitch zone) and the context of selected actions as text. Heat maps are derived from raw tracking data (Kernel Density Estimate) using the "ggplot2" package for the R statistical programming language. FB=Fullback; WM=Wing back. Taken from Ju et al. (unpublished).



^
Figure 5. The legend should read 'Figure 5. Team totals for the frequency of high-intensity actions during an English Premier League match using **(A)** 'blind' frequencies versus **(B)** contextualised frequencies related to tactical actions. These are based on the primary context but some hybrids are included. Taken from Ju et al. (unpublished).

Adding secondary options to the primary context starts to bring out extra directional and tactical granularity (Figure 6). For instance, in possession for the variable 'run with the ball' it can be seen that Team A does this primarily in central pitch zones (51%) or from wide to central zones (14%), while Team B who counter attacked regularly during the game primarily used the wide channels (60%) due to a high player density in central zones. This resulted in players on Team B 'moving to receive/exploit space' in the wide channel zones compared to Team A (35 v 12%) who dominated central zones. As expected, a higher proportion of 'run in behind' efforts for Team B occurred as the player moved from wide to more dangerous central zones than Team A (17% vs 7%). Once the ball has been advanced forward, intense 'support play' occurred and this was more common in central zones for Team A than B (76 vs 59%), and the opposite was true for the wide channel zones (14 vs 23%). Out of possession, recovery running composition was similar across teams but 'covering' highlighted the direct approach of Team B as Team A had to cover more during long passes (32%). From watching the game, Team A had a more coordinated 'pressing/closing down' strategy than Team B. Data also demonstrated that they would read the game more effectively and would hit high-intensity speeds and press before Team B players had even received the ball (54 vs 42%).

Figure 6. Secondary options for the primary contextualised variables during an English Premier League match using **(A)** in possession variables and **(B)** out of possession variables. Secondary options are depicted as a stacked donut chart and the frequency of actions are in the centre. These are based on the primary context but some hybrids are included. Taken from Ju et al. (unpublished).



TRANSLATION OF CONTEXTUALISED MATCH RUNNING PERFORMANCE: TRAINING

EXAMPLES FROM TWO ENGLISH PREMIER LEAGUE CLUB ACADEMIES

Recent research has not only revealed contextualised position-specific match running performance trends but has effectively translated these trends for the first time into individual and team positional training drills (Ade, 2019; Ade et al. 2016, 2020). This research demonstrated that high-intensity running distances were greatest for wide midfielders and lowest for centre backs with fullbacks, central midfielders and centre forwards falling somewhere in-between. However, as the data was contextualised it provided more insight into 'WHY' players produced purposeful tactical efforts in and out of possession. For instance, in possession, centre forwards carried out more high-intensity efforts in the offensive third of the pitch, whilst driving through the middle, running in behind, and breaking into the box. Whilst wide players like fullbacks and wide midfielders produced more high-intensity efforts overlapping and running the channel than other positions (Ade et al. 2016). They also performed more crosses after these runs than other positions due to more efforts finishing in wide attacking pitch areas. Out of possession, positions with a major defensive role in the team like centre backs, fullbacks and central midfielders produced more high-intensity efforts covering space or team-mates and recovery running whilst all positions performed frequent high-intensity efforts closing down the opposition.

The frequency, duration, distance, angle of turns of these contextualised efforts across positions are valuable prescription metrics when constructing combination or isolated drills, particularly when considered relative to one another (Ade, 2019; Ade et al. 2020). For a movement pattern, technical skill, combination play or tactical action to be included in the design of a position specific drill they adhered to one of the following criteria: (1) It occurs in >33% of efforts, (2) There is at least a small effect size difference compared to a minimum of two other positions, (3) In categories with a large number of variables (>3), there is a moderate standardised difference compared to the mean of the other variables. The third criteria allowed for actions that may not occur in a high percentage of efforts, but relative to the other variables are the most prominent and should therefore be included (e.g. heading for a centre back). Ade et al. (2016) reported the majority of high-intensity efforts do not include any ball contact (~60-75%), however for player enjoyment, technical skill development under fatigue and compliance such actions should be included.

The first drill designed used an appropriate blend of science gathered from the 'High Intensity Movement Programme' (Ade et al. 2016) and the art of coaching as evidenced by consultation with a UEFA Pro License football coach. This was a combination drill in which all positions are worked in unison with game- and position-specific ball work present. For effective drill design on a full-sized pitch, the start and end location of efforts were replicated to enhance the ecological validity of this drill, thus duplicating position-specific in and out of possession scenarios but with over-load (Ade, 2019). As speed endurance production and maintenance training typically induces sufficient metabolic overload (Ade et al. 2014) for aerobic and anaerobic adaptations in players (Iaia et al. 2015), this was the training mode used. This was used with English Premier League U17-18 Academy players (Figure 7A-C). The drill starts with the fullback producing an effort in the defensive third (first sequence) before overlapping the wide midfielder, to receive a pass in the wide attacking third to perform a cross (second sequence). Simultaneously, the centre forward breaks into the box to score while being tracked by the centre back, both having started in the middle third of the pitch (first sequence). The central midfielder drives through the middle of the pitch performing an arc run to support the attack ending with a possible shot on goal (second sequence). At the end, all positions produce a recovery run to individual pitch locations (final sequence). Using a speed endurance maintenance work to rest ratio of 1:2, all five positions produced 8 repetitions of ~30 seconds with 60 seconds recovery. This elicited an average and peak heart rate response of ~80 and 93% of maximal heart rate and produced a wide range of blood lactate concentrations following the final repetition of 6-16 mmol·L-1. Drills in which all positions are worked in unison with specific ball work adds variety to training while loading physical qualities alongside some tactical elements. More variation per rep is present in these circuits as the intensity can drop should one player perform a technical skill poorly (pass/touch) resulting in some positions having to slow down and alter their runs (Ade, 2019).

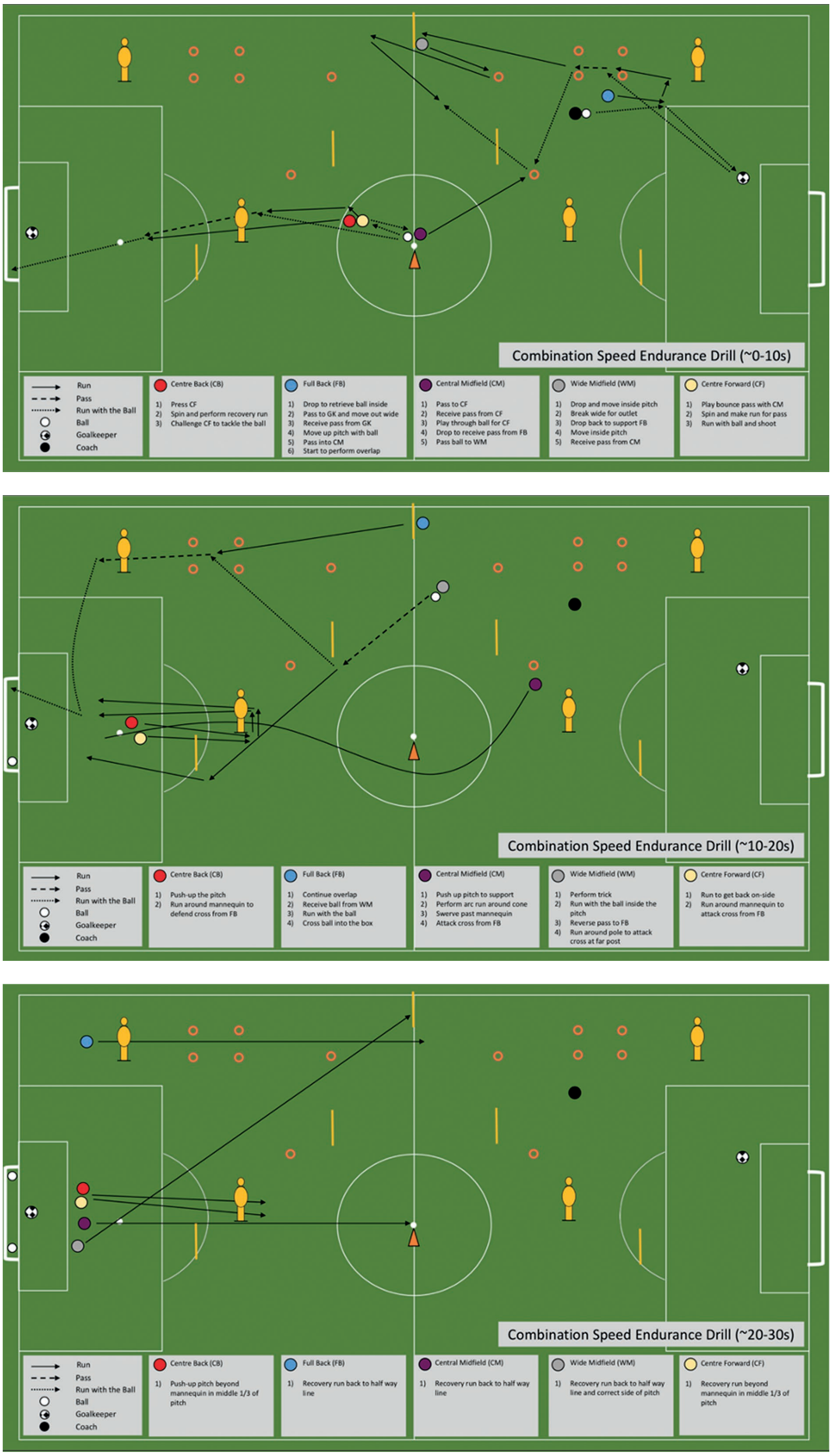
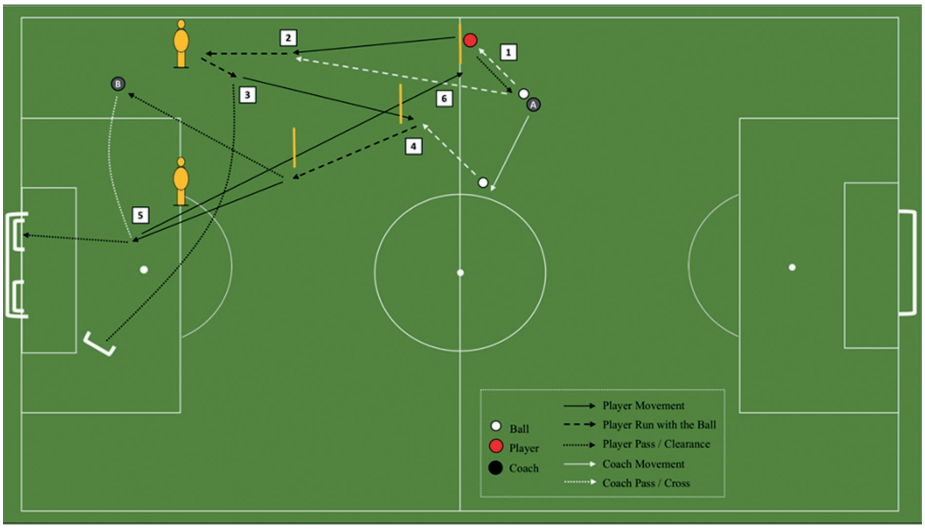


Figure 7. Position-Specific Speed Endurance Team Drill: **(A)** first sequence of drill: Coach plays ball inside FB to recover and play back to GK, at the same time the CM plays a bounce pass with CF before playing a ball over the top for the CF and CB to run on to contest. At the same time the WM drops to support the play but then pushes up and wide for an outlet for the GK. The FB then moves wide to receive the ball from the GK, CM drops to support the FB. The FB plays to the CM, the WM drops and moves inside the pitch to support the play. The CM passes to the WM whilst the FB performs an overlapping run. At the same time the CF and CB challenge for the ball over the top in a 1v1 situation resulting in the either the CF shooting on goal or the CB performing a clearance. **(B)** second sequence of drill: FB continues to perform overlapping run, CB pushes up the pitch whilst the CF performs a recovery run. The WM performs a trick upon receiving the ball from the CM, runs with the ball inside the pitch before playing a reverse pass out wide to the FB. The CM performs an arced run before driving through the middle of the pitch. The WM continues to run through the middle of the pitch. The CB and CF turn around the mannequin and start to accelerate into the box. The CM continues to drive through the middle of the pitch performing a swerve inside the mannequin. The FB runs with the ball and crosses into the box. The CF and CB run into the box to attack the ball whilst the CM and WM attack the front of the box and back post, respectively. **(C)** final sequence of drill: All players perform recovery runs back to set positions. See text above for description of drill. Taken from Ade (2019).

Given the high demands associated with the game, injuries of various severity are inevitable. Thus, the application of contextualised match data to rehabilitation drills can be used to ensure the stimulus is not just position specific but unique to the individual traits of a selected player and style of play. Drills using intervals enable practitioners to target certain physical qualities to ensure players adapt and thus return successfully. Similar to the above speed endurance drill, this requires a player to perform intense football activity for 20-30 seconds using recovery periods between 40-180 seconds, which is repeated 8-10 times dependent on the aim of the drill (production training has a work to rest ratio of ~1:5-1:6 and maintenance training has a work to rest ratio of ~1:1-1.3). This taxes players aerobically and anaerobically whilst involving the ball, so is ideal preparation before a return to team training (complementing drills emphasizing other qualities; Ade et al. 2020). Thus, contextualised patterns were translated into isolated position specific conditioning drills for players during end-stage rehabilitation. An example for a wide midfielder and fullback can be seen in Figures 8A-B, with some movements adapted to the team's tactical requirements for each position. GPS data captured during a speed endurance maintenance session (work to rest ratio of 1:2) completed by Liverpool FC Academy players returning from injury revealed that for selected speed thresholds (>14.4 and >19.8 km·h-1) wide midfielders (120 & 56 m) and fullbacks (104 & 60 m) covered greater distances per repetition across these drills than centre backs (68 & 16 m), central midfielders (93 & 28 m) and centre forwards (80 & 30 m), which is consistent with match trends (Bradley et al. 2009). Furthermore, centre backs and forwards covered the lowest overall distance per repetition (215 & 233 m·min-1, respectively) but performed greater total accelerations and decelerations (n=14 & 15) than full backs (n=11) and wide midfielders (n=9) though similar to central midfielders (n=13). High intensity accelerations and decelerations were more frequent for full backs (n=6), centre backs (n=6) and forwards (n=4) than central midfielders (n=3) and wide midfielders (n=3). This also elicited an average and peak heart rate response of ~85 and 91% of maximal heart rate and produced blood lactate concentrations following the final repetition of >14 mmol·L-1. This format can also be useful for 'top up' sessions when players are not getting game time and can be adapted to train multiple positions in unison to add a dynamic scenario. The conditioning coach not only prepares players for the demands necessary for training but also familiarises them with ball striking, discrete positional movements, orientation of space on the pitch, whilst providing a reactive stimulus so players are exposed to uncontrolled movements when returning to training with additional players. Elite clubs should use their analysis department to study player movements to create bespoke drills that are not only position specific but ideally individual specific (moving away from 'blind' distances/frequencies).

A



B

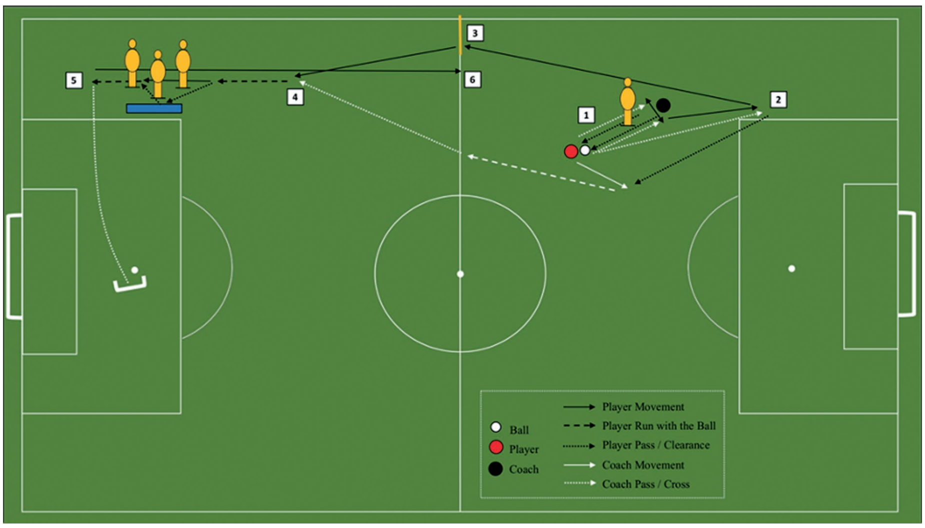


Figure 8. (A) End Stage Rehabilitation Drill for a Wide Midfielder. (1) play bounce pass with coach A and make a run down the channel. (2) receive pass from coach A, run with the ball, perform a trick in front of mannequin. (3) execute in-swinging cross into mini goal, then perform recovery run. (4) receive another pass from coach A, perform a trick and run with the ball driving inside the pitch before passing the ball wide to coach B. (5) break into the box to receive a cross from coach B and finish into mini goal. (6) perform recovery run back to original start position on half way line. Please note: players are given freedom for some decision making while the coach will vary the type of pass and cross e.g. players have the option to perform a trick and beat the mannequin during phase (2) to perform out-swinging cross into mini goal.12 **(B)**. End Stage Rehabilitation Drill for a Fullback. (1) Coach and FB play a one-two on either side of mannequin, moving FB side-to-side. (2) Coach plays ball down the inside for the FB to recover, FB sprints to recover the ball, turns and passes to coach inside the pitch. (3) FB overlaps coach around pole and receives pass in final third. (4) FB runs with ball and dribbles through mannequins. (5) FB delivers cross into mini goal. (6). Recovery run to the halfway line. Please note: players are given freedom for some decision making while the coach will vary the type of pass e.g. players have option to play off bounce board during phase (4) and cut back to play in swinging cross during phase (5).12 Individual player traits in terms of movements, tactical/technical events in training/games can also be added to conditioning drills for ecological validity purposes. Given the complexity involved in returning a player to training after injury, this drill is only one example from the players detailed end-stage rehabilitation plan. Taken from Ade et al. (2020).

CONCLUSION

Match running performance data are typically 'blind' but by adding context to the numbers it provides a much needed narrative and improves its translational power as a tool for insights and designing drills. The authors are not stating that 'blind' distance is not informative (WHAT), far from this but by combining it with context (HOW and WHY) enhances the metrics value. More work is still needed to get this type of analysis into a fit for purpose package and to reduce the time to insight for practitioners. Please be aware that some data are unpublished and just a snapshot of our industry case studies and thus not a scientific paper. As football is now in an information age, it is not just good enough to be able to 'know' but actually 'know how' to apply data into practice. This leap from data into practice is only really possible by blending science with the art of coaching and this piece provides numerous examples of using match analysis to design training drills. The use of integrated data is imperative to this translation process and it would be advantageous for more work of this nature to be disseminated to identify areas of best practice and to drive innovation in data-practice applications.

REFERENCES

Ade JD. Speed Endurance in Training in Elite Youth Soccer Players. Doctoral thesis, Liverpool John Moores University. 2019. Available from: http://researchonline.ljmu.ac.uk/id/eprint/11390/1/2019AdePhD.pdf

Ade JD, Drust B, Morgan O, & Bradley PS. Physiological Characteristics and Acute Fatigue Associated with Position Specific Speed Endurance Soccer Drills: Production vs Maintenance. *Sci & Med Football*. 2020; DOI: 10.1080/24733938.2020.1789202

Ade JD, Fitzpatrick J, & Bradley PS. High-intensity efforts in elite soccer matches and associated movement patterns, technical skills and tactical actions. Information for position-specific training drills. *J Sports Sci*. 2017; 34 (24), 2205-2214.

Ade JD, Harley JA, Bradley PS. Physiological response, time-motion characteristics, and reproducibility of various speed-endurance drills in elite youth soccer players: small-sided games versus generic running. *Int J Sports Physiol Perform*. 2014;9(3):471–479.

Akenhead R, Hayes PR, Thompson KG, French D. Diminutions of acceleration and deceleration output during professional football match play. *J Sci Med Sport*. 2013;16(6):556–561.

Bangsbo J, Norregaard L, Thorso F. Activity profile of competition soccer. *Can J Sport Sci* 1991; 16:110-116.

Barnes C, Archer DT, Hogg B, Bush M, Bradley PS. The evolution of physical and technical performance parameters in the English Premier League. *Int J Sports Med* 2014; 35: 1095-1100.

Bradley PS, Ade JD. Are current physical match performance metrics in elite soccer fit for purpose or is the adoption of an integrated approach needed? *Int J Sports Physiol Perform*. 2018; 13: 656–664.

Bradley PS, Archer DT, Hogg B, Schuth G, Bush M, Carling C, & Barnes C. Tier-specific evolution of match performance characteristics in the English Premier League: it's getting tougher at the top. *J Sports Sci*. 2016;34(10):980–987.

Bradley PS, Evans M, Laws A, Ade JD. 'Context is King' when interpreting match physical performances. *Football Med & Sci Magazine*. 2018; 24:42–45.

Bradley PS & Scott D. Physical Analysis of the FIFA Women's World Cup France 2019™. Zurich: FIFA. Available from: https://img.fifa.com/image/upload/zijqly4oednqa5gffgaz.pdf. 2020; 1-168. [Accessed July 2020].

Bradley PS, Sheldon W, Wooster B, Olsen P, Boanas P, & Krustrup P. High-intensity running in English Premier League soccer matches. *J Sports Sci*. 2009; 27, (2), 159-168.

Bush, MD, Archer, DT, Hogg, R, and Bradley, PS. Factors influencing physical and technical variability in the English Premier League. *Int J Sports Physiol Perform*. 2015; 10: 865–872.

Bush M, Barnes C, Archer DT, Hogg B, Bradley PS. Evolution of match performance parameters for various playing positions in the English Premier League. *Hum Mov Sci*. 2015; 39:1–11.

Carling C, Bradley PS, McCall, Dupont G. Match-to-match variability in high-speed running activity in a professional soccer team. *J Sports Sci*. 2016;34(24):2215-2223.

Collet C. The possession game? A comparative analysis of ball retention and team success in European and international football, 2007-2010. *J Sports Sci* 2013; 31: 123-136.

Da Mota GR, Thiengo CR, Gimenes SV, Bradley PS. The effects of ball possession status on physical and technical indicators during the 2014 FIFA World Cup Finals. *J Sports Sci* 2016; 34: 493-500.

Di Mascio M, Bradley PS. Evaluation of the most intense high-intensity running period in English FA premier league soccer matches. *J Strength Cond* Res 2013; 27: 909-915.

Di Salvo V, Gregson W, Atkinson G, Tordoff P, Drust B. Analysis of high intensity activity in Premier League soccer. *Int J Sports Med* 2009; 30:205-212.

Gregson W, Drust B, Atkinson G, & Di Salvo V. Match-to-match variability of high-speed activities in premier league soccer. *Int J Sports Med*, 2010; 31, 237–242.

Iaia FM, Fiorenza M, Perri E, Alberti G, Millet GP, Bangsbo J. The Effect of Two Speed Endurance Training Regimes on Performance of Soccer Players. *PLoS One* 2015; 10 (9): e0138096.

Martín-García A, Casamichana D, Díaz AG, Cos F, & Gabbett TJ. Positional Differences in the Most Demanding Passages of Play in Football Competition. *J Sports Sci & Med*. 2018a; 17(4), 563-570.

Martín-García A, Gómez AD, Bradley PS, Morera F, & Casamichana D. Quantification of a Professional Football Team's External Load Using a Microcycle Structure. J *Strength & Cond* Res. 2018b; 32(12):3511-3518.

Mohr M, Krustrup P, Bangsbo J. Match performance of high-standard soccer players with special reference to development of fatigue. *J Sports Sci* 2003; 21:519-528.

Osgnach C, Poser S, Bernardini R, Rinaldo R, di Prampero PE. Energy cost and metabolic power in elite soccer: a new match analysis approach. *Med Sci Sports Exerc*. 2010; 42:170–178

Reilly T, Thomas V. A motion-analysis of work-rate in different positional roles in professional football match-play. *J Hum Mov Stud* 1976; 2: 87–97.

Gain an understanding of the science behind football tactics and the use of new technologies to enhance the game.

CERTIFICATE IN FOOTBALL TACTICAL ANALYST

REGISTER NOW

CERTIFICATE IN TECHNOLOGY AND SCIENCE APPLIED TO SPORT

REGISTER NOW

BARÇA COACH ACADEMY

REGISTER NOW

BARÇA
INNOVATION HUB