Computers and Electrical Engineering 81 (2020) 106505

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Secrets of soccer: Neural network flows and game performance☆

Massimo Marchiori[a,b,*], Marco de Vecchi[b]

[a] European Institute for Science, Media and Democracy, Belgium
[b] University of Padua, Italy

ARTICLE INFO

Article history:
Received 10 December 2018
Revised 29 October 2019
Accepted 4 November 2019
Available online 19 November 2019

Keywords:
Soccer
Football
Games
Sport
Networks
Flows
Efficiency

ABSTRACT

Soccer is the most popular sport in the world, with currently over three billion fans. There are various reasons for this success, but a unique feature of soccer stands out: every match has a high level of unpredictability. For instance, it is not uncommon for a less skilled team to sometimes defeat a better team. Moreover, a match that is apparently decided in favor of a team can suddenly change course, even with few minutes left, ending with a completely opposite result. These highly dramatic effects have brought popularity to this sport, making every match a complex event where the outcome is far from granted. This same complexity has however challenged all attempts of data analysis: the “secrets of soccer”, that is to say the recipes for success, are still an unknown realm, defying all common statistical approaches. In this study we try to shed some light on these secrets by introducing a novel approach that uses neural network flows. We transform a team play into a corresponding brain-like structure, an abstraction that we analyze using measures of efficiency, assessing the “quality of thinking” of the brain. This way, we can view any soccer match as an alternate battle of minds and explore how far this parallelism can help to solve some fundamental open problems, like finding an effective recipe for success, and establishing the best field control strategies.

1. Introduction

The game of soccer has in the latest years gained an enormous traction, becoming widespread in almost all parts of the world. For example, soccer is the number one sport for global audience, tv viewers, internet popularity, number of professionals, market value and more. Its success is due to several factors, but also partly to its unpredictability: lovers of this sport praise the fact that soccer is fascinating because of its lack of apparent determinism. Within a single game, having a better team does not guarantee success, and even teams belonging to inferior leagues can sometimes beat world famous top teams. Moreover, even within a single game playing better does not guarantee success: a team can be vastly superior as far as all common statistics (like ball possession, shoots on target and so on) are concerned, and then eventually still lose the game. An extreme case of this paradox occurred for instance in the English Premier League, when a team of underdogs (Leicester City Football Club) won the 2016 title: Leicester won many matches even if very often puzzlingly behind in most

☆ This paper is for CAEE special section SI-csc.Reviews processed and recommended for publication to the Editor-in-Chief by Guest Editor Dr. Xiaokang Zhou.

* Corresponding author.
E-mail addresses: massimo.marchiori@unipd.it, massimo@math.unipd.it (M. Marchiori).

https://doi.org/10.1016/j.compeleceng.2019.106505
0045-7906/

game statistics. This uncertainty in games has brought soccer to an exceptional level of popularity, having a situation where each game, despite the level of strategy put in by managers and players, still have some surprise factor.

Along with the success of soccer, there have also been interest in trying to study the complexity of the game and unravel some of the strategies that influence the final game score. However, even with the increasing progress in technology and data science techniques, the secrets of this game have still proved hard to unlock. There have been many works related to soccer analysis using various tools (see for instance [1–5]), studies trying to identify the various kinds of team formation [6], as well as studies performing general statistical analysis so to help understand the mysteries of this sport [7]. However, despite all the attempts to define and analyze significant metrics for soccer analysis (see e.g. Clemente et al. [8]) or to find some sense about statistical data on matches [9], soccer has remained a baffling quest for data analysis, despite the ongoing availability of big data sets (cf. Rein and Memmert [10]). In fact, most commonly used statistics to define a team performance (like total shots, shots on target, shots off target, ball possession, number of off-sides, fouls received, corners and so on) did not manage to effectively grasp what makes a team win a match (cf. Clemente et al. [8]). This complexity is in fact part of the fascinating aspects of this game: whatever the reasons that eventually make a team win a single match, they are neither simple nor obvious.

In this paper, which is a revised and expanded version of [11], we tackle the problem of soccer analysis using a different approach. The key point is to shift perspective, starting from a fundamental question: what is the focus aspect to consider in a soccer game? The most natural answer would be the players: the players are the prime actors in the game, and they determine success or failure. And so for instance, having better players looks like an obvious recipe for success. Following this intuition, works in the literature have been trying to study individual player behavior, attempting to grasp how such behavior can actually impact the game (see for instance [12–15]). However, as said earlier, these approaches have not managed to provide definitive answers, failing to find meaningful metrics for success. In this work we instead change perspective, and do not consider the individual soccer players as the main actors of the game. Instead, we view players as subordinate to another component: the play field. The soccer field itself is considered as the primary component, and players are seen as functional objects that enable *information exchange* among the various zones of the field. In doing so, we turn each match into a kind of brain-like network, where ball passing becomes an instance of a neural network flow, and the whole entity "playfield plus players plus ball passing" becomes the equivalent of a so-called *Team Brain*. This way we can change point of view, and instead of looking at a match as a soccer battle, we view it as a corresponding "battle of minds": whoever thinks better wins.

In this study we introduce the approach, and then test it on real data from an international soccer tournament. The obtained results show that this parallelism seems to work and can help to provide valuable insights to unlock at least some of the complexity of the game. This allows to go beyond classic statistical approaches (which proved to be unsuitable for the game), and to start answering the fundamental question that is the key to this sport: what does it really make a team win or lose a single soccer match? We show how the quality (efficiency) of the corresponding Team Brain looks like the secret key to win a game, correlating qualitatively with actual win-draw-lose results. Moreover, the Team Brain model also provides quantitative information about the final score, enabling to guess with high probability the result of game.

Extending the analysis, we also investigate the Team Brain dynamics with respect to time. In order to win do we need a Team Brain with a few intense moments of thoughts, or is it instead better a steady thinking flow? In other words, is the recipe to success given by peaks of brilliant play, or is it more important to be consistent all along the match?

Another important problem we tackle is in-game prediction power: how much can we infer the final result also in mid-game, without waiting for the complete end-of-game data?

Last but not least, we then also proceed to see whether there are areas of the playfield that are more important than others, and correspondingly find some surprising results about the perception of a Team Brain. We show that the "mind" composed by a Team Brain views the play field using different geospatial metrics, and this discrepancy from our classical perception could also explain the previous difficulties in identifying proper significant statistics for soccer. Relatedly, we also focus on the actual physical size of the neural zones and investigate what are the most meaningful areas to consider in order to build a winning Team Brain.

Answering all these questions provides interesting food for thought and helps to unravel some of the secrets of this complex sport.

The paper is organized as follows. In Section 2 we introduce the concept of Team Brain, transforming the match of a team into a network brain-like connection structure. In Section 3 we then introduce the machinery that allows us to define how efficient such a structure can be, which is the prelude to the next Section 4, where we then leverage a team match into the new concept of "battle of minds". In Section 5 we apply this approach to real game data, showing how the performance of the Team Brain is a key metric correlated to the final result of a game. In Section 6 we delve deeper, and see whether we can actually get a result predictor for a game. In Section 7 we study the impact of time on the battles of mind, and investigate its relationship with goals and game progression. We then proceed in Section 9 analyzing whether there are zones of the field that are more critical than others for the success of a team, obtaining rather surprising results that revisit concepts like midfield, forward, wings and so on. Section 10 then critically analyses the concept of neural zone with respect to size, trying to determine the meaningful physical size range of the neural zones. In Section 11 we briefly hint at some possible future lines of research expanding on the current analysis. Finally, Section 12 ends the paper summarizing the results.

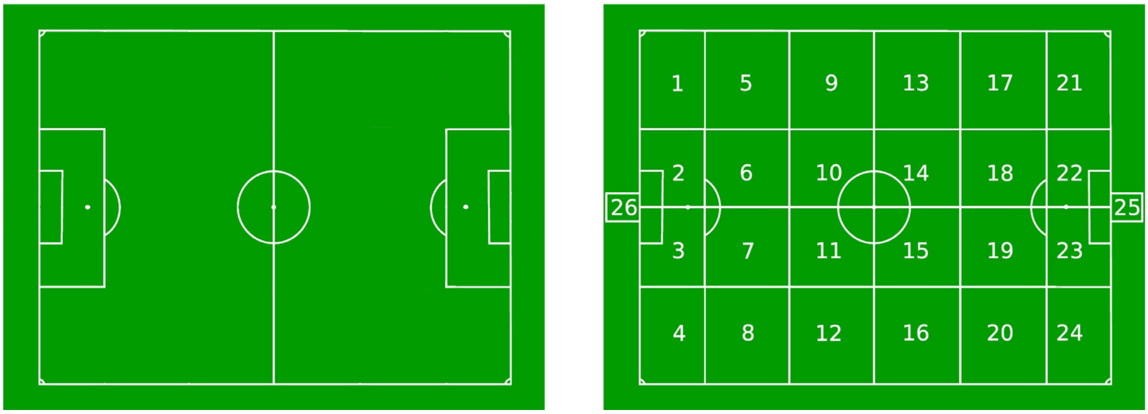


Fig. 1. The Soccer Field (left) separated into neural zones (right).

2. The team brain

In this Section we introduce the concept of Team Brain and show how to turn a soccer match into a corresponding neural structure. As hinted before, we reverse the classic approach of considering the team players as first-class actors, abstracting from them and focusing instead on the playfield as primary factor, seeing players as components of a bigger cyber-structure. In particular, we consider team players as functional to one basic action: information exchange. In this interpretation of the game, "information" is brought along by the ball. So, ball passing can be seen as an exchange of information not among players, but instead among zones of the playfield. In this model, using brain terminology, players can in fact be interpreted as "neurons" performing information exchanges among the various brain (field) zones. So, this view abstracts from individuals, and focuses instead on team behavior: players are worth for their contribution to ball exchange within the field structure.

The first thing to do in order to define such cyber-structure is to specify what are the corresponding "neural zones", given that the "neurons" (the players) are not still objects, but instead they move around the playfield. We therefore abstract from player movements and define still areas of the playfield as main building blocks, thus dividing the soccer field into corresponding "neural zones" that are in fact the first-class objects to consider in a soccer game.

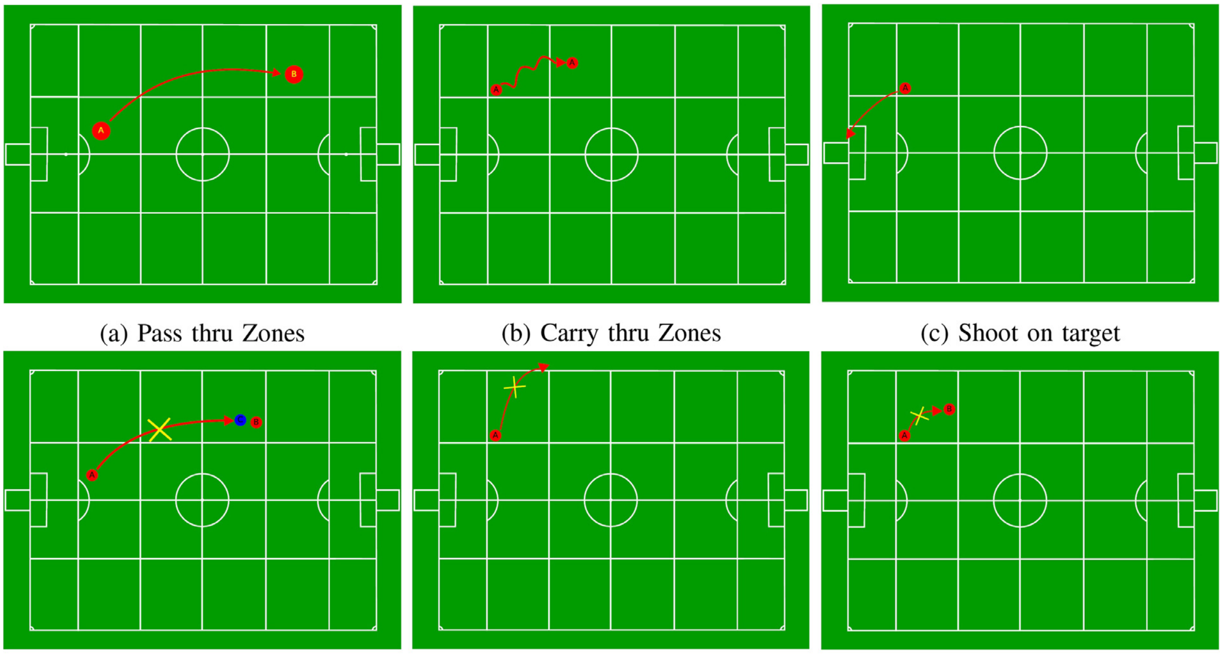
This neural zones division is illustrated in Fig. 1: we divide the soccer field into 26 zones (see later Section 10 for different divisions), following a grid structure. The first 24 zones are a grid-like division of the soccer field, and the last two (25 and 26) indicate the special zones made by the goals. Note that the penalty areas correspond to zones (2,3) and (22,23). We do not create special zones indicating the goal areas, opting for a uniform distribution of the zones in the field (see however Section 10 for an analysis of the zones size and the corresponding implications).

Ball movement then defines connections between these zones (equivalent to links in the corresponding network), according to various possibilities that are listed in the following.

In the simplest case, indicated in Fig. 2a, player A passes the ball to another player B of the same team, thus creating a directed connection between the corresponding source and target zones. Other ball movements are possible, and each corresponds to a different information transition.

Another case to consider is when the player itself carries the ball running from one zone to an adjacent one. In this case (see for example Fig. 2b), there is also a directed connection from the start zone to the target zone: the only difference is in the speed of the ball. Using the abstraction of information exchange, this lower speed implies a less efficient information transition when compared to the first case. In other words, passes are more efficient way to transmit information. We consider this difference by differently weighting runs and passes: the links created by passes have more weight than those created by runs. This means that weights correspond to a *speed* measure (how quickly information is transferred), or alternatively they can be seen as a *capacity* measure (how big the information channel is).

The final case to consider for link creation deals with the special zones 25 and 26: these areas are "reached" when a player of the opposite side shoots on target, see Fig. 2c. There are then other three remaining possibilities for ball movement, and all three do not contribute to the information exchange, and therefore do not create links. The first case is when a pass is intercepted by an opponent player (see Fig. 2d): given the pass is not successfully completed, it is discarded. Note that an alternative approach would be to further distinguish whether the pass was fruitful to the opponent team, so counting this as a pass in the other team brain: given this would also imply a qualitative subjective distinction among passes, we opted for a simpler approach without ambiguities. The second case is when a pass goes out of the field (see Fig. 2e): this also corresponds to an unsuccessful pass and is therefore discarded. Finally, the last possible case is when the ball is brought along or passed within the same zone (Fig. 2f): local ball movements of this kind are also discarded, as they do not contribute to the information transfer among zones.



(a) Pass thru Zones (b) Carry thru Zones (c) Shoot on target

(d) Pass Intercept (e) Out of field (f) Inside Zone

Fig. 2. Ball Movements and field grid.

We have seen how the actions of players on the ball (note we do not care about the other movements of the players without ball) can cause an information transfer between two field zones. This transfer can then be represented with a *directional weighted link*. Since there can be more links between the same zones, what we obtain when turning play time into this representation is actually a *multigraph*, that is to say a graph where there can be multiple links between two vertices. Using the parallelism with a brain structure, zones can be seen as neurons, and the links as synapses: so, neurons connected one another with more synapses exchange information more efficiently than those connected with less synapses.

Given a team within a soccer match we call *Team Brain* the corresponding multigraph obtained using the above rules. Note that a Team Brain can actually be constructed from every temporal subset of a match (so for instance, allowing to measure local changes during the match): we first consider full-match Team Brains, in order to see how their global behavior impact the actual match performance (Sections 5, 6), and then go on to see how smaller time spans affect the Team Brain and its relationship to the match result (Section 7).

3. Measuring team brains

In the previous section we have seen how to turn the game play of a team into a Team Brain, represented by a multigraph network. The next step is to define a measure telling us how well this brain can "reason", following the underlying idea that better Team Brains would then correspond to winning teams. Given that the Team Brain is modeled with the idea of neural zones that exchange information, we use the notion of *network efficiency* introduced in [16,17]. As the name suggests, network efficiency is (roughly speaking) a measure of how efficient it is the information exchange within a network. This notion has been successfully used in a variety of fields, including analysis of brain structures (see for example Avena-Koenigsberger et al. [18]). We proceed by stating its basic definition, and then go on with its application to Team Brains.

3.1. Network efficiency

We being by introducing some basic definitions of graph theory, and then provide the formal definition of network efficiency. A (*simple*) graph G is a pair (V, E) where $V = \{u_1, u_2, \dots, u_n\}$, $|V| = n$ is a finite set of vertices and $E \subseteq V \times V$, $E = \{(u_i, u_j),\ i \neq j\}$, $|E| = m$ is the set of edges that links couples of nodes. These graphs are called *topological*. A graph G can be represented by a $n \times n$ adjacency matrix A with entries $a_{ij} = 1$ when $(u_i,\ u_j) \in E$, and $a_{ij} = 0$ otherwise. $a_{ii} = 1$ denotes self loops. A weighted graph is defined as $G = (V, E, w)$ where w is a function that assigns real values to edges.

Metrical graphs (also known as spatial graphs) extend weighted graphs as they are spatially embedded: that is to say, every node exists within a space endowed with Euclidean coordinates. More specifically, $G = (V, E, C, w)$ where $C =$

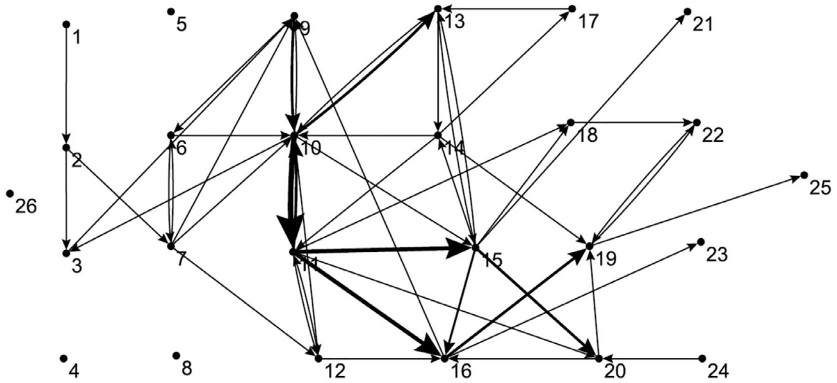


Fig. 3. Fifteen minutes of the Spanish Team Brain (2008 European Championship Spain vs Sweden).

$\{(x_1, y_1), (x_2, y_2) \cdots, (x_n, y_n)\}$ is the set of node coordinates (e.g., a spatial position in terms of latitude and longitude) and the function w might assigns, for instance, Euclidean distances between nodes.

The *global efficiency* of a graph G [16] can then be defined as the average of efficiencies ϵ_{ij} between two vertices of the graph in this way:

$$E_{glob}(G) = \frac{\sum_{i \neq j \in V} \epsilon_{ij}}{n(n-1)} = \frac{1}{n(n-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}}$$

Here, the efficiency ϵ_{ij} is inversely proportional to the "distance" d_{ij} between the two vertices i and j. The notion of distance that is used depends on the specific knowledge about the system. For example, d_{ij} can have different meanings in unweighted and weighted networks: in the first (unweighted) case it could be defined as the number of hops between two nodes in the shortest path (a choice giving rise to the so-called *topological efficiency*), whereas in the second (weighted) case it could be defined as the sum of all edge distances in the shortest path (a choice known as *metrical efficiency*). In general, global efficiency is a convenient universal measure also because it applies equally well both to connected and to unconnected networks (assuming that unconnected vertices have an infinite distance).

Efficiency can be applied also to weighted graphs and to multigraphs, such as the current case of the Team Brain. First, every weighted graph (where the weight correspond to capacity of the channel) can be turned into a corresponding metrical graph by dividing the original distance between two vertices by the corresponding weight of the edge. Secondly, a multigraph can then be turned into a weighted graph by summing up the weights of each multiple link, thus obtaining a single link with the sum of the weights.

The numerical range of global efficiency, as defined above, ranges from 0 to $+\infty$, but in practical applications it is often more convenient to normalize it, so to deal with a limited range of values. The classic normalization procedure for efficiency is to divide the unnormalized value defined above by the so-called *ideal network* K_n, so obtaining a normalized version of efficiency defined as $E^{norm}_{glob} = E^{glob}(G)/E_{glob}(K_n)$. Here, the ideal network is defined as the most efficient network that is possible within the given context. The classic definition of K_n (cf. Latora and Marchiori [19]) takes as ideal network of a given network the corresponding fully connected graph (therefore obtaining a graph with $n * (n-1)/2$ vertices). The vertices in the ideal network are then connected in the fastest way allowed by the given context (the given graph). In the specific case of a metrical multigraph (as the one we are dealing with) this means that the distance between every pair of vertices in the ideal weighted graph is given by their physical distance on the field divided by the maximum weight of the original weighted graph.

As an example, consider for instance the graph in Fig. 3, where the link with maximum weight is the one between node 10 and node 11: the ideal network in that case is the fully connected graph where the distance between every pair of vertices is their physical distance divided by that maximum weight.

This normalization procedure leads to a range for normalized efficiency of $0 \leq E^{norm}_{glob}(G) \leq 1$ (so, the value 0 corresponds to the least efficiency network, and then efficiency can grow until the maximum value of 1, corresponding to the most efficient network). For better readability, in the following (unless stated otherwise) we will use the term efficiency to indicate the normalized global efficiency.

3.2. Team brain efficiency

Having seen how network efficiency can also be applied to multigraphs, we can then use it to measure the behavior of Team Brains. In order to do so, we simply turn each Team Brain multigraph into a corresponding metrical graph: the final distance (d_{ij}) between two areas i and j is obtained dividing the metrical distance between the two areas (calculated for instance point-to-point from the center of every area) by the sum of all the weights of the links from i to j.

For instance, the Spanish Team Brain corresponding to 15 min (for better graphical visibility) of the 2008 match Spain vs. Sweden is shown in Fig. 3.

We can then calculate for every Team Brain its corresponding global efficiency (roughly speaking, indicating how well the Team "thinks"), and use that measure to compare Team Brains, therefore passing from a soccer game to a "battle of minds".

An important thing to notice is that the number of passes could be different for each team, therefore risking privileging the team with higher ball possession: in other words, the efficiency of a Team Brain could just be a by-product of ball possession. In fact this is not the case because we use efficiency, which is normalized with respect to the best network: the latter (remember the definition of normalization stated in 3.1) is based on the maximum weight (maximum number of passes). This means that having a network with more links (a Team Brain with more connections) does not necessarily imply a bigger normalized efficiency: the unnormalized efficiency (E_{glob}) gets bigger, but on the other hand the efficiency of the ideal network might also get bigger, depending on the maximum number of passes among zones. In practice, this means that even if a team collects more passes than the other one, this does not necessarily imply a better efficiency, and in fact it could well be quite the opposite, depending on the distribution of the passes. The normalization process of efficiency therefore makes the analysis not just dependent on the number of passes of a team, and so makes the analysis different from ball possession (which is anyway a statistic that, as said earlier, does not manage to capture the complexity of soccer).

4. Team performance and battles of minds

The alternate view of a soccer match as a "battle of minds" allows to then look for correlations between these two worlds.

However, when dealing with the soccer world, we have first to define what level of detail we are interested in. For instance, we might be interested either in the final score of the game (a very precise outcome), or in a lower level of detail, like for instance just knowing who the winner is, independently on the number of scored goals. We investigate both choices, defining the corresponding two ways of expressing the performance of a team in a match.

As said, one way to define the performance of a team is to consider the result, comprehensive of the scored goals. At this level of detail, a win by 3 to 0 (say) is superior to a win by 1 to 0. We therefore define the *delta performance* (within a set soccer match) as the difference between the number of goals scored by a team and the number of goals scored by the opponent. The delta performance is therefore an integer that can assume both negative and positive values.

The second way to define the performance of a team is to abstract from the precise number of goals, and just consider the result: win, draw, or loss. We call this the *win performance*, and model it as a three-valued range +1, 0 or -1, representing respectively win, draw or loss.

We can then explore the relationship between battles of soccer (soccer matches), using either definitions of performance, and battle of minds (Team Brains). In the Team Brain view, a battle is a comparison of efficiencies: in order to obtain again a single score, we can just consider the relative efficiency of a Team Brain versus another one as the ratio of their efficiencies.

5. Correlation results

Having defined how to turn a soccer match into a battle of minds leads to the next step: is there a correlation between these two views of the game? In other words, is the concept of a Battle of Minds (with its related concept of *better thinking* given by the measure of network efficiency) truly representative of the game result? Can we semantically compress a complex game dynamic into this abstraction, therefore obtaining insights on the fundamental forces leading to winning or losing a game?

In order to test this hypothesis, we considered the matches of the (final phase of the) European Championship 2008 (made up by 16 teams, and 31 matches overall). The choice of this competition was made because of the reasonable number of tournament matches (versus a whole national championship, for instance), and the availability of all the recorded matches by one of the authors. We then performed via video play (and replay) a visual analysis of the matches and corresponding manual creation of the ball transitioning dataset. Note that the dataset was manually compiled, given the present lack of publicly available soccer datasets having accurate ball passing information (see the discussion in Section 11), which for the moment made hard to actually also consider other championships.

From the dataset we then extracted the corresponding Team Brain multigraphs, and then proceeded with the calculations of the "battles of minds" and correlation to game performances (both delta and win performance).

The correlation for the case of delta performance turned out to be 0.80, whereas the correlation for the case of win performance 0.75. The first thing to note is the very high level of correlation in both cases, which clearly shows that the efficiency of the Team Brain is a surprisingly good indicator of how well the team performs on the field. The second thing to note is that in fact when abstracting information (passing from delta to win performance) the correlation lowers: an indication that Team Brain efficiency is apparently a more precise quantitative indicator, and not only a qualitative indicator, for the final result of a team.

6. Prediction results

Apart from correlation, we might also want to investigate the relationship between battles of minds and the actual result of a match in a different way. Correlation tells us there is a general correspondence between these values, but like every average value there might be also cases where the correspondence is not perfect (and so, for few individual matches, the battle of minds might not perfectly coincide with the outcome of a match). So, in addition to a global correlation measure, we can also focus on individual matches, and try to see what is the local (for a single match) prediction power of the battle of minds versus the outcome.

However, in soccer games (unlike some other games) the outcome is three-valued, given that each team can not only win or lose, but also draw. Ideally we would like to check whether via the Team Brain concept we manage to distinguish among all these three results, but there is a potential problem: when mapping game performance to real numbers (in our case, the [0,1] range of efficiency) how do we define a draw result? Imposing exact numeric equivalence seems unfeasible, given that efficiency is after all an estimation.

Given we are comparing two numbers, we can go back to a two-valued outcome by testing for “weak prediction”, defined as predicting whether a team will *win or draw*. In other words, weak prediction can tell us whether a team is generally superior, in that at least it will not lose versus the other. In terms of battle of minds, this is equivalent to test whether any time a team has a better ($\geq$) efficiency than another then the same time either wins or at least draws. Checking for weak prediction on the dataset reveals that the battle of minds has 96.77% accuracy: every match successfully passes weak prediction, apart from one, the championship final (Germany vs. Spain, result 0 to 1), where the battle of minds is in favor of Germany (1.49 ratio). Overall, we can see this 96.77% successful result as a further confirmation of the intuition behind the battle of minds, in that it provides not only a global (correlation) correspondence (both for winning and delta performance), but also weak predictive power for individual matches. Failure in the championship final can be accounted either for the very special status of that match, where maybe also other additional factors (extreme psychological pressure?) contribute, or to the general complexity of soccer itself, that makes soccer games not yet fully explainable.

What if we want to go beyond weak predictions, and see whether the draw result can somehow be grasped? That is, what about exploring “full prediction”, where we try to deal with the full three-valued outcome of a game, and so see whether from a single battle of minds we manage to also grasp the eventual draw result? In fact, having weak predictive power as seen above might not necessarily imply any kind of predictive power for the draw result itself. In order to have predictive power for the draw result we can for instance employ thresholds: check whether the efficiency values for draws are separated from the efficiency values for wins. If we have separation, we know that (given that weak prediction works, as seen above), we can eventually define a function using thresholds that can separate draws from wins, and so get full predictive power.

In fact, analysis of the data reveals that, apart from the aforementioned championship final, there is a threshold that can successfully separate all draws from wins, given that for all matches the highest draw value of a Team Brain (1.55) is lower than the lowest win value (1.84). So, we can for instance set a threshold to 1.6, and obtain a win-or-draw full predictor, defined as

- If the efficiency of the Team Brains of Team A and B in a match is respectively $E(A)$ and $E(B)$, then
- if $E(A) > E(B)$ and $E(A) > 1.6$ then A wins
- if $E(B) > E(A)$ and $E(B) > 1.6$ then B wins
- otherwhise, A and B draw

which, again, has 96.77% accuracy. Note that the mismatched brain battle value in the championship final (1.49) stays within the draw zone, thus somehow showing this unique mismatch was not that deep as to indicate victory instead of defeat.

7. Global versus local time

So far, we have correlated the outcome of battles of minds with the game result. The same definition of battles of minds relies on calculating the overall efficiency of the brain-like structure generated during the whole game. One might wonder whether in fact there are also local variations of efficiency that matter during the match. In other words, is the overall result the by-product of a “thinking process” that has to take into account the whole match, or can we also focus on smaller time scales (parts of the match)? We delve into this question by first focusing on goals, and then on game progression.

7.1. Goals and bursts

Goals are the climax of every soccer match, and they determine the result. So, it is interesting to see whether something special happens in their time proximity: in other words, is there significant change in the thinking of the mind of a team before a goal? Can it be that a goal is just the by-product of a better time-local thinking of a team during the match? If this is the case, then the overall correlation and prediction results that we have obtained so far could have a further explanation: goals might be provoked by “bursts” of faster local thinking (increase in efficiency), causing a goal to be scored. So, a team might be thinking normally, and then some “thinking burst” might occur, causing a goal to be scored.

Table 1
Game progression and effectiveness of the battles of minds (delta / win).

Time	15 min	30 min	Half Time	15 min (2^{nd} half)	30 min (2nd half)	End Game
Correlations	0.23 / 0.27	0.38 / 0.44	0.48 / 0.44	0.62 / 0.54	0.77 / 0.71	0.80 / 0.75
Precision	28.8% / 36%	47.5% / 58.7%	64% / 55%	77.5% / 72%	96.3% / 94.7%	100% / 100%
Linear gain	+17.3% / +21.6%	+14.3% / +17.6%	+12.8% / +11%	+11.6% / +10.8%	+11.6% / +11.4%	0% / 0%

We tested this hypothesis by analyzing for each team the fractions of a match that occur before scoring a goal (using a time span of five minutes), and comparing them with the other parts of the match (averaging the other remaining five minutes parts), so to compare the "speed of thought" before a goal versus the average thinking behavior during the match. More formally, the time division is done as follows. If $t_1, \ldots, t_n$ are the times at which a team scored its n goals, we can define the time periods $[t_1 - 5\text{ m}], \ldots [t_n - 5\text{ m}]$, where $5m$ stands for 5 min, as being the "pre-goal time slots" (in case $t_1 < 5$ m we discard the period, being too short for comparison). We then divide all the remaining game time into 5 min interval (the "no-goal time slots") by starting from the beginning of each pre-goal time slot ($t_i - 5$ m) and going backwards ($[t_i - 10\text{ m}, t_i - 5\text{ m}], [t_i - 15\text{ m}, t_i - 10\text{ m}], \ldots,$ always disregarding a time period if its length is less than 5 min. In order to cover also time after the last goal t_n we add a fictitious goal time t_{n+1} coinciding with the final time of the match.

So for instance if a team scored two times at 12 m02 s (12 min and 2 s) and at 61 m23 s, and the match ended at 91 m56 s, we consider the pre-goal time slots to be [7 m02 s, 12 m02 s] and [56 m23 s, 61 m23 s]. The remaining time is divided into 5 min slots as described above, obtaining the no-goal time slots [2 m02 s, 7 m02 s], [16 ms23 s, 21 m23 s], ..., [51 m23 s, 56 m23 s], [61 m56 s − 66 m56 s], ..., [86 m56 s, 91 m56 s].

The results, somehow surprisingly, showed that there seem to be no special behavior of a team before a goal: in fact, the efficiency before a goal is fluctuating, sometimes bigger sometimes smaller than the average: correlating the pairs (efficiency of a pre-goal time slot, average efficiency of the non-goal time slots in the same match) for all the matches we obtain −0.16.

We did the same test by also tackling the dual case, that is analyzing the fractions of a match before a goal scored by the other team, obtaining similar results (fluctuating behavior, and correlation 0.13)

Summing up, the overall result of a match seems more influenced by the global thinking behavior of a team, rather than by isolated bursts of faster (or slower) thinking.

7.2. Game progression

In Sections 5 and 6 we have investigated the various relationship between battles of minds and game performances. Localizing time near to goal events, as seen in the previous subsection, does not provide significant result, emphasizing the importance of global long-term thinking versus short-term local bursts. However, how critical it is to consider the whole match, versus a part of it? For instance, we might question whether half times are already enough to say something about the result, or if on the contrary we do need the whole match thinking behavior in order to get high correlation and prediction power. We investigated this aspect by analyzing *game progression*, that is calculating the battles of brains not only at the end of the match, but instead progressively during the match. This also allows us to see whether games are somehow already determined after a certain period (given the prior behavior of the teams), or if instead they stay uncertain until the very end.

The obtained results are summarized in Table 1, where we show the progressive behavior of the battle of minds using six time divisions The first row shows the correlations values for delta and win performances, whereas the second row (precision) shows how precise the results are w.r.t. the ones obtained by the full end of game analysis. The third row (linear gain) shows the difference between the actual precision of the analysis versus a linear precision progression (so, showing how fast the augment in precision is).

As we can see, the first interesting thing to note is that there is always positive correlation, starting from the first span of 15 min into a match: so, even a small beginning of the match can already provide some (weak) indication on the outcome.

The second thing to note is that the precision of the analysis then grows monotonically with time, therefore providing a better and better correlation and prediction for the result, giving significant values of correlations already at half time. In fact, the level of precision obtained after 30 min in the second half is so high (96.3% and 94.7% for delta and win performances respectively) to make the last part of the game almost superfluous: in other words, the outcome of a match seems to be dictated by the behavior in the first 75 min.

Last but not least, it is interesting to see that precision does not proceed proportionally with time, as shown by the linear gain row: the first 15 min provide a considerable gain (+17.3% and +21.6%), which then progressively decreases at 30 min and at half time, until 15 min in the second half where it then stabilizes (always in a super-linear gain) until the end of the game.

8. Underlying field structure

We have previously stressed the fact that the basic constitutive layer of the Team Brain is made by the physical soccer field. The correlation results between Team Brain efficiency and performance then tell us that what really matters is the ball movement, that is the equivalent of a transmission exchange (a “thought” in brain terminology). This information exchange (ball passing) occurs within a geometrical environment (the soccer field), with its notion of distance.

We might wonder how much this physical structure, with its actual distances, is a factor for game success. In other words, what happens when for instance we drop distances, passing from a metrical space to a topological one (in other words, imposing that all distances are equal)? Dropping distances allows us to verify whether ball passing is heavily depending on the underlying physical structure, or all it matters is actual ball circulation. We therefore proceeded to calculate correlations in this topological case, obtaining new values of correlations: 0.76 for delta performance, and 0.70 for win performance. So, this correlation drop shows that the underlying metrical structure does in fact play a role, and taking it away means having a Team Brain structure that is less representative of the final result of a match.

9. Critical areas and field perception

Given that distances do matter, we might go further, and ask an even more challenging question. Does the Team Brain perceive the field in exactly the way we perceive it? In other words, the Team Brain interacts with the underlying field structure, and dropping its metrical information (as seen previously) degrades performances. But the metric information we have provided to the Team Brain is the plain one given by field geometry: Euclidean distances on the field. It might be that the Team Brain senses the field differently, using a different geometry. So, for instance, some areas of the field might be perceived as more important than others, and correspondingly the geometry of the field be different.

These kinds of questions can be answered by modifying the metrical structure of the playfield and observing the effects on the Team Brain (and related correlations to performance). In other words, just like taking away good metrical information leads to worse correlation, other metrical modifications leading to better correlations could be the indication that the Team Brain senses the field in a different way than the ordinary.

We therefore proceeded to analyze whether there are critical areas or not: in other words, if some areas of the field are sensed as more important than others, and so the corresponding perceived geometry of the field changes from the perspective of the Team Brain. Fig. 4 shows the seven main scenarios that we tested when looking for critical areas, and corresponding different perceptions.

For instance, the first scenario (Fig. 4a) checks for differences between the wings (highlighted in red) and the rest of the field. What we do is to alter the distances of ball passes ending in a wing side and observe the effect on the Team Brain efficiency correlations. Note that altering distance can also be seen, equivalently, as changing capacity, given the correspondence between these two concepts that are combined together in the final abstract distance (cf. Section 3).

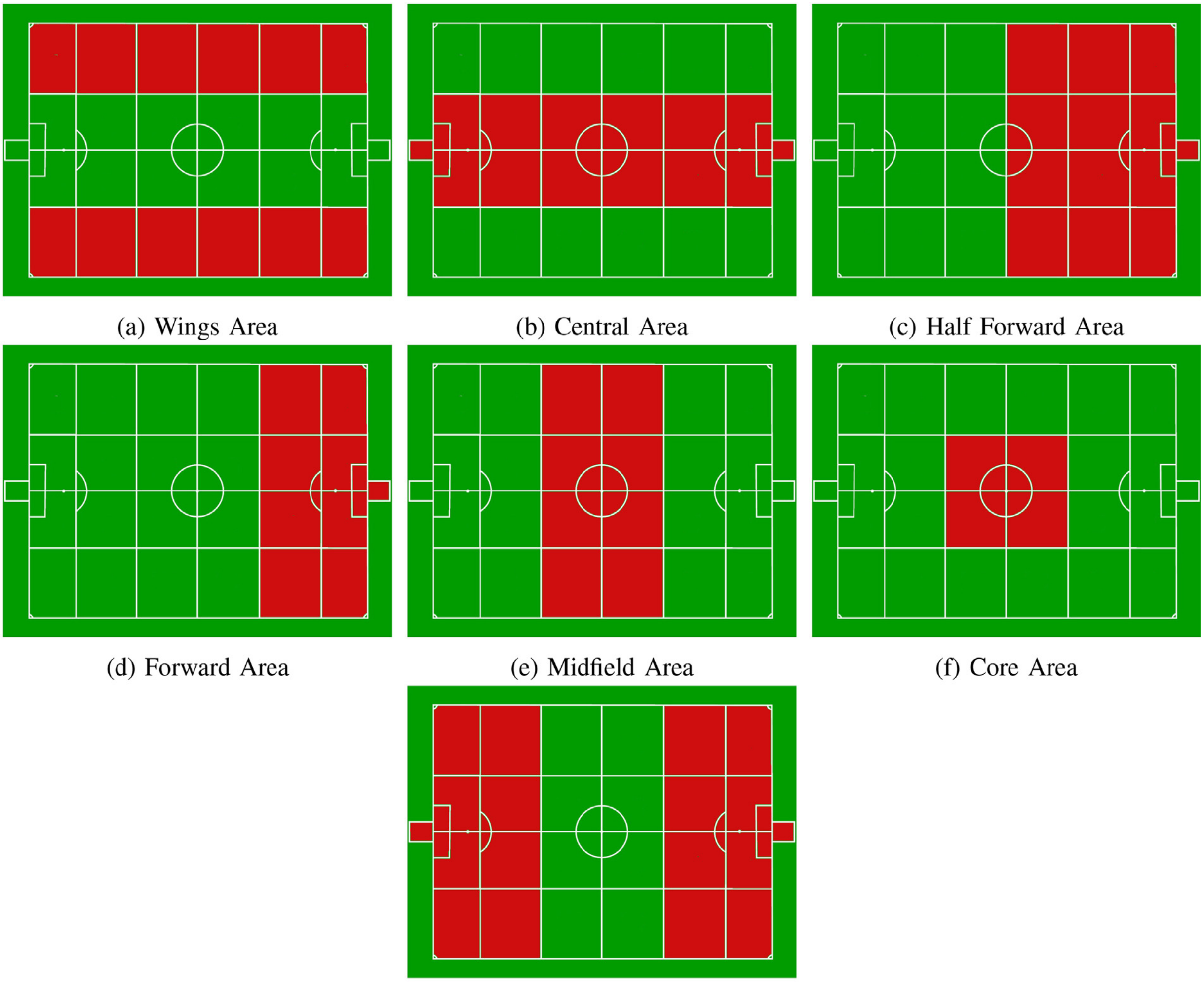
We do the same for other scenarios, like the dual one (central stripe area, Fig. 4b), two scenarios checking on the importance of “forward” in soccer (half forward area, Fig. 4c, and forward area, Fig. 4d), two scenarios checking for the importance of the midfield areas (Fig. 4e for the classic midfield, and Fig. 4f for the core central area within midfield), and the scenario dual to midfield (the “extremes”, Fig. 4g).

For each scenario, we tried various modifications, ranging from diminishing distances (up to a 0.4 multiplicative factor) to stretching them (up to a 3 multiplicative factor), both for the length (“x” axis) and the width (“y” axis). We hereby summarize the results, proceeding from the first scenario to the last one:

- **Wings area:** modifications didn’t produce much changes in correlations, obtaining in fact worse results, with the exception of distance stretched by 2 in the x axis (which produced a slightly higher correlation of 0.83)
- **Central area:** results got worse, except for distance stretched around the 0.5 value in the y axis (corresponding correlation value of 0.90)
- **Half-Forward area:** worse results, apart from the case of distance stretched by 2.0 in the x axis (0.81)
- **Forward Area:** same results as Half-Forward
- **Midfield Area:** same results as Half-Forward, but with correlation value of 0.91
- **Core Area:** worse results
- **Extremes Area:** same results as Half-Forward

These results can give significant insights for actual game strategy. First, they say that some zones that are currently highly valued in soccer, like wings, half-forward and forward are not more important than the other zones (!). In other words, rather paradoxically, pushing the ball “forward” is not necessarily good for winning the game: the concept of “forward” is not better than ball passing in the other zones, making up for a balanced Team Brain structure. The same for the concept of wings or core: forcing gameplay along the wings or the field core is not better than using the other zones, and so usage of the wings and core areas is equivalent in importance to the other parts of the field.

The other results concern the Central and the Midfield area. The results show that the midfield area has a bigger perceived importance along the x axis (in other word, it is crucial to quickly pass midfield along the direction of length), whereas the central area has a lesser importance when passed along the y axis. But even more interestingly, these two scenarios are geometrically symmetrical, and have the best correlation value that is geometrically dual: 0.5 y stretch for Central, and 2.0 x stretch for Midfield. The corresponding perceived geometries would then be very close to each other, which brings us to a further hypothesis: what happens if we use these stretching values, and use them uniformly all over the field? In other words, what happens when we try to combine these two scenarios into one, and apply the stretching to the whole field?



(a) Wings Area (b) Central Area (c) Half Forward Area

(d) Forward Area (e) Midfield Area (f) Core Area

(g) Extremes Area

Fig. 4. Critical Areas of the playfield.

We then applied these modifications to the whole field and obtained the following results: correlations got worse when we shrank the field along the x axis, whereas they increased when we enlarged it, reaching a peak around the 2.0 stretch, with a rather impressive correlation value of 0.92 (the highest so far obtained!). This means that the two single scenarios (Central and Midfield) can be seen in fact as partial views of the single unified scenario, and corresponding geometry: the Team Brain senses the field *as twice longer* compared to our "normal" senses (see Fig. 5). So in a sense, it is two times more difficult to move the ball along the direction of length, and the corresponding perceived geometry changes. Interestingly, this top level of correlation (0.92) stays the same both in the delta performance and in the win performance case, making the perceived geometry nicely sweep away the differences between quantitative and qualitative correlations to performance.

10. Neural zones size

In order to extract a brain-like network structure from the playfield, we have divided it into 26 zones. Apart from the two special zones corresponding to the goals, one might wonder whether this division is the only possible one. While it is an intuitively reasonable division, allowing us to distinguish between center and wing areas (in the horizontal axis), and allowing to distinguish in each midfield between forward and backward zones, the question may arise of whether there are other possible divisions that are still meaningful. In other words, the current division proves that the neural zones, as defined, are meaningful, that is to say the physical area they represent do play a role, and constitute a kind of "building block" for the game. But what about smaller, or bigger areas? What happens when we change scale, and pass to finer (or larger) divisions? Are the physical areas that the division generates still meaningful? In other words, with the original neural

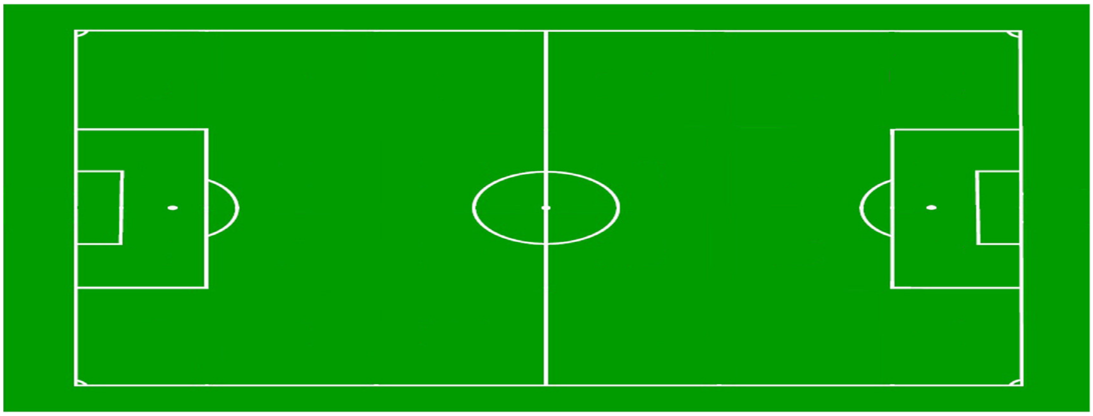


Fig. 5. The Soccer Field geometry as perceived by the Team Brain.

Table 2
Grid divisions and correlations to delta / win performances.

	x:3	x:4	x:5	x:6	x:7	x:8	x:9
y:2	0.13 / 0.09	0.40 / 0.35	0.51 / 0.45	0.56 / 0.51	0.48 / 0.43	0.37 / 0.33	0.21 / 0.15
y:3	0.28 / 0.23	0.66 / 0.62	0.73 / 0.70	0.76 / 0.72	0.70 / 0.68	0.56 / 0.50	0.39 / 0.33
y:4	0.32 / 0.24	0.67 / 0.62	0.79 / 0.76	**0.80 / 0.75**	0.74 / 0.71	0.61 / 0.54	0.44 / 0.36
y:5	0.24 / 0.19	0.61 / 0.56	0.65 / 0.58	0.66 / 0.61	0.63 / 0.59	0.53 / 0.48	0.36 / 0.31
y:6	0.11 / 0.07	0.38 / 0.34	0.49 / 0.43	0.51 / 0.47	0.47 / 0.42	0.33 / 0.30	0.17 / 0.13

zone division we have imposed a certain scale factor, but we still don't know what happens when we change scale, given that other scales could provide better or worse results (cf. Marchiori and Possamai [20]).

Stating the same problem from the game perspective: when we talk about zones to control in the playfield, and that need to be properly connected in order to win, can we say something more about the actual size of the zones? What is the best granularity in a strategy of field control and ball passing that a team should focus on?

In order to investigate this issue, we can proceed by using different sizes for the neural zones, using other grid divisions. Given the current grid division is 4×6 (playfield width divided by 4, and length divided by 6) we can for instance experiment with wider zones (playfield width divided by 1–3, and length divided by 1–5) as well as with smaller zones. However, while experimenting with other grid divisions is apparently not a problem, there is a problem we must face when dealing with the data set. The current data set has been collected by using the 4×8 grid structure, and as such positioning of the ball movement has been recorded as relative to the default neural zones. Therefore, we currently lack more precise positional information, and as such the only alternate grids that we could use are the ones "subsumed" by the current 4x6 grid, that is to say the grids 4×3, 4×2, 4×1, 2×6, 2×3, 2×2, 2×1, 1×6, 1×3 and 1×2 (omitting the trivial grid 1×1). So, we could investigate larger neural zones, but not smaller ones.

In order to overcome this problem, we proceed to simulate a finer positioning of the ball: for each ball position recorded within a certain neural zone, we proceed to calculate a random position within the same neural zone. This way we can simulate a dataset with precise positioning (under the assumption that the positions within a neural zone are almost equally distributed), and consequently use also other grid divisions, so to fully explore the issue of the more meaningful sizes of the neural zones.

Table 2 shows the result for various grid combinations, corresponding to divisions of the "x" length axis from 3 to 9, and divisions of the "y" width axis from 2 to 6, showing the corresponding levels of correlations with delta performance and win performance.

The obtained data show various interesting things. First, the grid division 4x5 gets almost equal precision as the base 4x6 one. Second, the values tend to get worse as far as we go further away from the 4x5 and 4x6 grids, both with larger and with smaller neural zones. Third, this decrease in the correlations is not linear w.r.t. the differences in the grid size: whereas the grid sizes close to 4x5 and 4x6 still maintain somehow high correlations, the values rapidly degenerate when further deviating from those sizes.

So all in all, this is an indication that in fact there are some more meaningful sizes for the neural zones, corresponding to the 4x5 and 4x6 grid divisions. Other sizes close to these ones are still meaningful, but their significance rapidly diminishes the moment we go to smaller or larger sizes.

Anecdotally, an interesting thing to note is that the number of zones in the two best cases (4×5 and 4×6) is respectively 20 and 24, whereas the combined number of players is 22: so, there might be some kind of correlation between the number of players involved in the game, and the consequent distribution of the critical zones to control in the field: in this respect, it would be interesting to test this hypothesis with other sports (like for instance basketball) having similar ball movement principles but a different number of players.

11. Future work

In order to further expand on the current analysis, the key factor to future work is collecting more data with enough level of precision: data about match results are widespread, whereas data about ball passes are available commercially but not academically. In fact, such larger datasets could automatically be collected using computer vision techniques for ball tracking (see for example Kamble et al. [21]) allowing not only to better validate the results presented here but also to reason at a larger scale. For instance, we could focus on what happens to a single team going beyond a single game: how does the Team Brain of a specific team change during the season, or versus the same opponents in different matches? Similarly, we could apply this kind of analysis to specific sets of teams (correlation of large number of matches rather than single matches). Another interesting idea worth exploring is to verify the potential of in-game progression prediction, as per Subsection 7.2, with respect to betting systems (cf. Rue and Salvesen [22], Fitt et al. [23], Croxson and James Reade [24]).

Larger automated datasets might also make sense of the current anomalies (think of the championship final), as well as allowing to apply this approach to other sports that are always based on ball passing, like for instance basketball [25], volleyball, cricket and so on, seeing whether the same neural approach can be used, and what similarities or differences eventually exist. This would allow to look for fundamental principles that can be applied to unravel the complexity of many different sports.

12. Conclusions

In this study we have introduced a new model of soccer analysis, based on the key intuition that players are not first-class entities, but are rather parts of a primary entity (the soccer field), and are functional to the common goal of information exchanges among its various zones. The soccer field becomes the equivalent of a brain container, and the players play the role of neurons, thus acting within a combined cyber-entity by activating neural network flows. This way we are able to turn soccer games into “battles of minds”, and see whether this change of perspective is able to actually grasp the key for success within a single match. The obtained results show that this approach is extremely promising, providing a high level of correlation between the efficiency of a Team Brain structure and the final game results, as well as predictive capabilities. We have also shown how the quality of thinking of a Team Brain is more important, from the time perspective, at global level than at local level: the recipe for success seems to be having a Team Brain that has a steady good functioning, rather than relying on sudden “bursts” of better thinking to score a goal. Also, the dynamic evolution of the Team Brain structure during the game shows that the actual building process is monotonically progressive and give more and more information about the final result. This information augmentation follows a super-linear growth that tends to make the last part of the game less important: in other words, surprises in the last part of the game might in fact just be the product of the quality of thinking that had been accumulated during the match, and not instead a special time zone somehow detached from the rest of the match.

We have also performed an analysis of the critical areas of the soccer field, that somehow revisited some concepts given for granted in the soccer community. This brought to quite surprising results, showing that some common sense in soccer (like the importance of “forward”, of core midfield, of wings play) might be in fact not accurate at all: the real important thing seems to be the right balancing of passes among the various areas, producing an overall more efficient brain-like structure. These counter-intuitive results might also explain the failure of common statistics like ball possession, shoots, corners and so on, to grasp the ultimate secret for winning a soccer game. We have also seen how the actual perception of the soccer field by the Team Brain is likely to be different than the normal physical one, producing a kind of “alternate reality” where the field is perceived using a different geometry, actually stretched two times.

Finally, we also have devoted some attention to the scaling granularity of the neural zones, showing that the most meaningful sizes truly defining the concept of a “zone of interest” in soccer are given by 4x6 or 4x5 grids, and so are in a range of 20–24. These are the zones that need to be considered by a team when looking to produce an efficient Team Brain, and consequently win the game.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Halvorsen P, Saegrov S, Mortensen A, Kristensen DKC, Eichhorn A, Stenhaug M, et al. Bagadus: An Integrated System for Arena Sports Analytics: A Soccer Case Study. In: Proceedings of the 4th ACM multimedia systems conference. MMSys '13. New York, NY, USA: ACM; 2013. p. 48–59. http://doi.acm.org/10.1145/2483977.2483982.

[2] Perin C, Vuillemot R, Fekete JD. Soccerstories: a kick-off for visual soccer analysis. IEEE Trans Vis Comput Graph 2013;19(12):2506–15. doi:10.1109/TVCG.2013.192.

[3] Rehman A, Saba T. Features extraction for soccer video semantic analysis: current achievements and remaining issues. Artif Intell Rev 2014;41(3):451–61. https://link.springer.com/article/10.1007/s10462-012-9319-1.
[4] Janetzko H, Sacha D, Stein M, Schreck T, Keim DA, Deussen O. Feature-driven visual analytics of soccer data. In: 2014 IEEE conference on visual analytics science and technology (VAST); 2014. p. 13–22. doi:10.1109/VAST.2014.7042477.
[5] Schlipsing M, Salmen J, Tschentscher M, Igel C. Adaptive pattern recognition in real-time video-based soccer analysis. J Real-Time Image Process 2017;13(2):345–61. https://link.springer.com/article/10.1007/s11554-014-0406-1.
[6] Bialkowski A, Lucey P, Carr P, Yue Y, Sridharan S, Matthews I. Large-scale analysis of soccer matches using spatiotemporal tracking data. In: 2014 IEEE international conference on data mining; 2014. p. 725–30. doi:10.1109/ICDM.2014.133.
[7] Shafizadeh M, Taylor M, Peñas CL. Performance consistency of international soccer teams in euro 2012: a time series analysis. J Hum Kinet 2013;38:213–26. https://www.degruyter.com/view/j/hukin.2013.38.issue-1/hukin-2013-0061/hukin-2013-0061.xml.
[8] Clemente FM, Sequeiros JB, Correia AFPP, Silva FGM, Martins FML. Computational metrics for soccer analysis: connecting the dots. 1st ed. Springer; 2017.
[9] Castellano J, Casamichana D, Lago C. The use of match statistics that discriminate between successful and unsuccessful soccer teams. J Hum Kinet 2012;31:137–47. https://www.degruyter.com/view/j/hukin.2012.31.issue--1/v10078-012-0015-7/v10078-012-0015-7.xml.
[10] Rein R, Memmert D. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. Springerplus 2016;5:1410. https://doi.org/10.1186/s40064-016-3108-2.
[11] Marchiori M, de Vecchi M. The team brain: Soccer analysis and battles of minds. In: Proceedings of the IEEE cyber science and technology congress (CyberSciTech 2018). IEEE; 2018.
[12] Gregson W, Drust B, Atkinson G, Salvo VD. Match-to-match variability of high-speed activities in premier league soccer. Int J Sports Med 2010;31(4):237–42. http://europepmc.org/abstract/med/20157871.
[13] Lago-Peñas C, Lago-Ballesteros J. Game location and team quality effects on performance profiles in professional soccer. J Sports Sci Med 2011;10(3):465–71. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3737821/.
[14] Dellal A, Chamari K, Wong DP, Ahmaidi S, Keller D, Barros R, et al. Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. Eur J Sport Sci 2011;11(1):51–9. https://doi.org/10.1080/17461391.2010.481334.
[15] Vogelbein M, Nopp S, Hökelmann A. Defensive transition in soccer – are prompt possession regains a measure of success? A quantitative analysis of German Fußball-Bundesliga 2010/2011. J Sports Sci 2014;32(11):1076–83. https://shapeamerica.tandfonline.com/doi/abs/10.1080/02640414.2013.879671.
[16] Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett 2001;87(19):198701.
[17] Marchiori M, Latora V. Harmony in the small-world. Phys A 2000;285(3–4):539–46.
[18] Avena-Koenigsberger A, Misic B, Sporns O. Communication dynamics in complex brain networks. Nat Rev Neurosci 2018;19(1):17–33. doi:10.1038/nrn.2017.149.
[19] Latora V, Marchiori M. Economic small-world behavior in weighted networks. Eur Phys J B 2003;32(2):249–63.
[20] Marchiori M, Possamai L. Micro-macro analysis of complex networks. PLoS One 2015;10(1):e0116670.
[21] Kamble PR, Keskar AG, Bhurchandi KM. Ball tracking in sports: a survey. Artif Intell Rev 2017. https://doi.org/10.1007/s10462-017-9582-2.
[22] Rue H, Salvesen O. Prediction and retrospective analysis of soccer matches in a League. J Royal Stat Soc 2000;49(3):399–418. https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9884.00243.
[23] Fitt AD, Howls CJ, Kabelka M. Valuation of soccer spread bets. J Oper Res Soc 2006;57(8):975–85. https://doi.org/10.1057/palgrave.jors.2602062.
[24] Croxson K, James Reade J. Information and efficiency: goal arrival in soccer betting. Econ J 2014;124(575):62–91. https://academic.oup.com/ej/article/124/575/62/5076978.
[25] Chen H-T, Tien M-C, Chen Y-W, Tsai W-J, Lee S-Y. Physics-based ball tracking and 3D trajectory reconstruction with applications to shooting location estimation in basketball video. J Vis Commun Image Represent 2009;20(3):204–16. http://www.sciencedirect.com/science/article/pii/S104732030800117X.

Massimo Marchiori is Professor at UNIPD (Italy) and CTO of EISMD (Belgium). At MIT/W3C (USA) he led the development of several world standards. He created Hypersearch (Google's forerunner), Volunia (the world-first social search engine), Negapedia (the negative version of Wikipedia). He won several awards, including the IBM research award, the Microsoft Data Science Award, the MIT TR35 award.

Marco de Vecchi, data analyst, graduated in Computer Science at the University of Padua (Italy) under the supervision of Professor Massimo Marchiori.