



twenty3  
SPORTS INTELLIGENCE

# Detection of Playing Style of Football Teams and Players using Latent Dirichlet Allocation

David Perdomo Meza

# Motivation

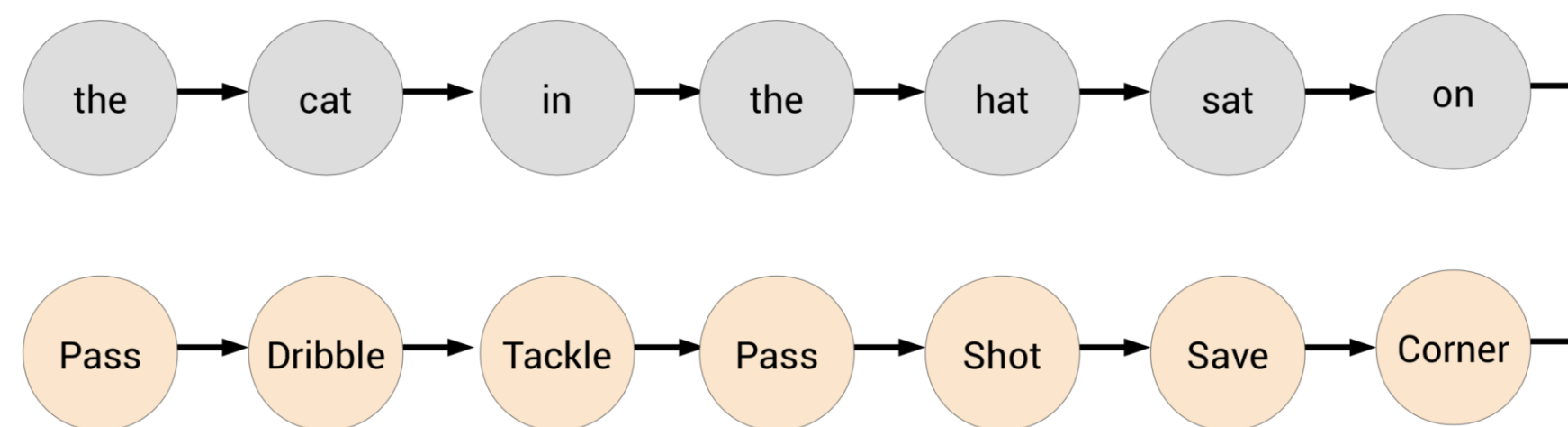
- **Objective:** Operationalising “playing style” of football teams and players



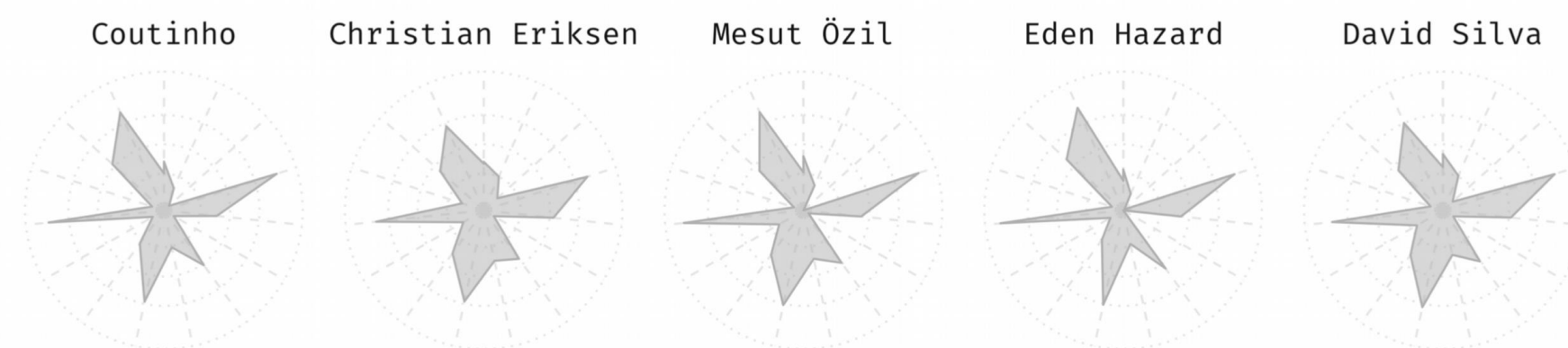
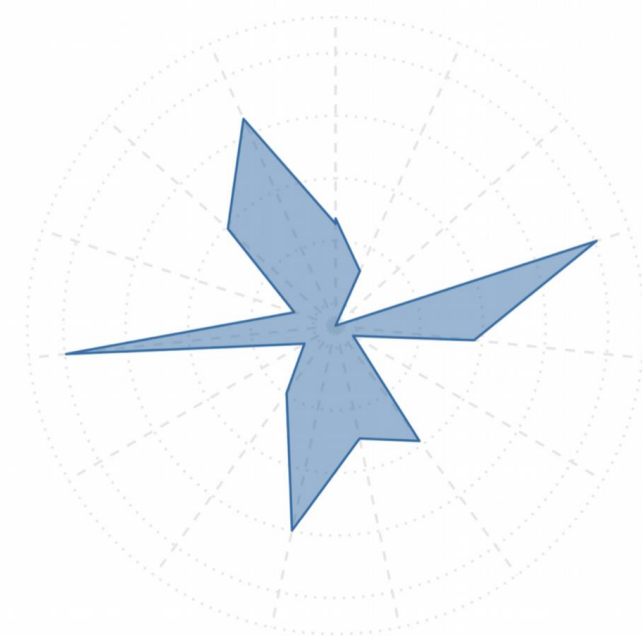
- **Supervised Research in Football:**
  - Low Scoring Sport
  - Aggregated Over Large Timeframes
  - Non-Descriptive
- Manual Analysis
- The question of style is **unsupervised**

# “Playing Style” as an Encoded Vector

- Vectorial representation of style is convenient and usable
- **EXAMPLE:** Ben Torvaney’s player2vec (word2vec from NLP)



Lionel Messi

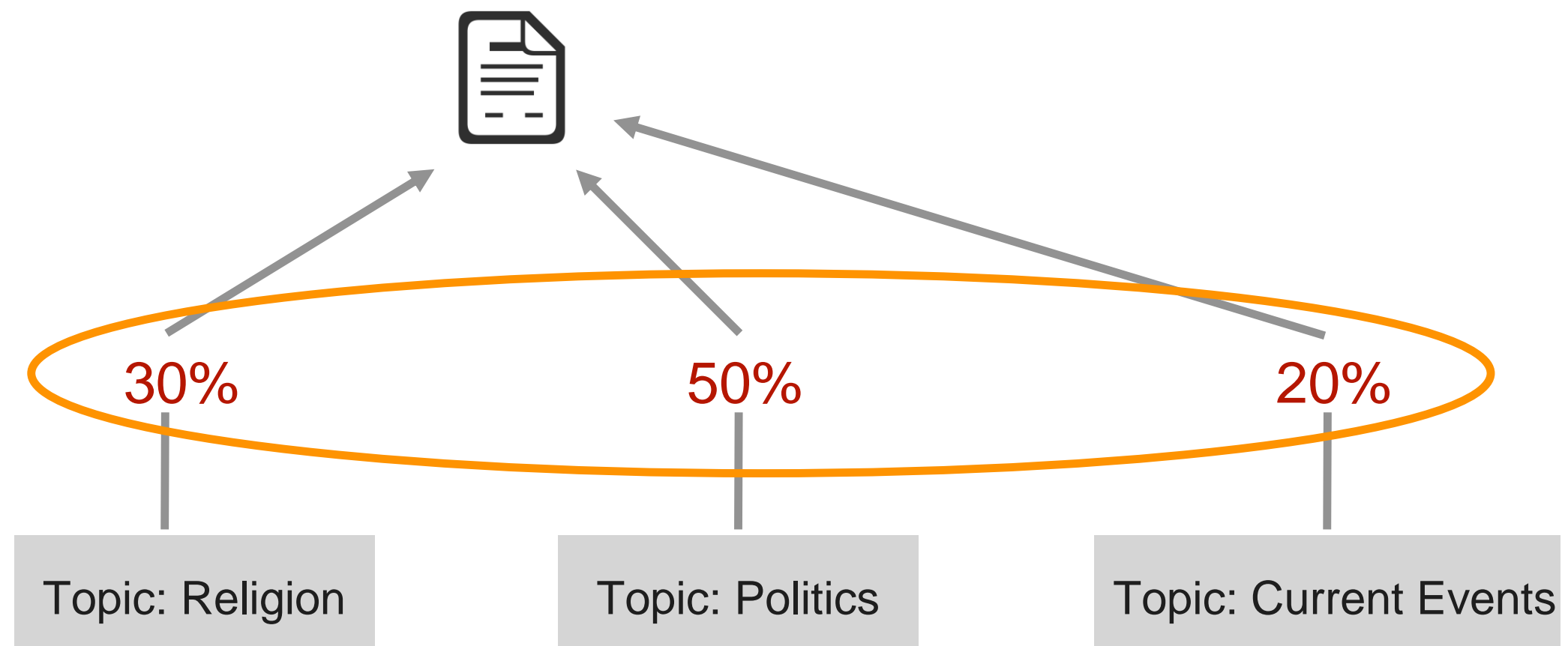


$$\begin{array}{ccccccc}
 \text{cat} & - & \text{dog} & < & \text{cat} & - & \text{phone} \\
 [0.2, \dots] & & [0.1, \dots] & & [0.2, \dots] & & [0.8, \dots]
 \end{array}$$

- Stylistic? (Cautious) Yes
- How do we know?
- Problems:
  - Batch Execution
  - “Player first” rather than “style first”
  - Non-descriptive - doesn’t correspond with intuitive notions of style

# Topic Extraction

- Non-Batch Execution
- Descriptive: Needs to correspond to intuitive notions
- Topic Mixture Models: Vectorisations!



- Unsupervised: **Latent Dirichlet Allocation**

# Eureka Moment: !

“Just how different words determine the topics of a document, different in-play actions determine the style of a team/player”

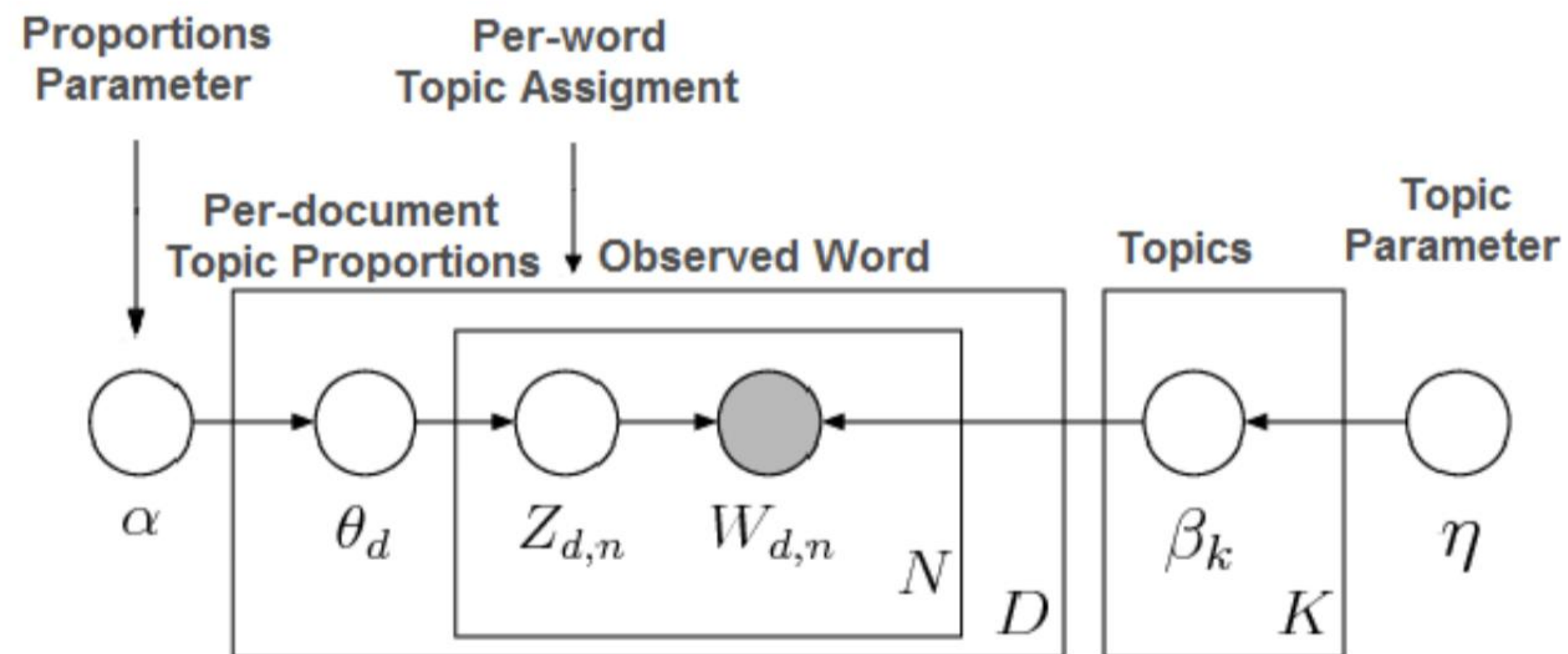
	god	president	law	mail	research	people	christ	university	women	memory	government
Doc 1	3	1	3	1	4	3	1	1	1	0	1
Doc 2	1	2	1	0	4	4	0	4	3	2	2
Doc 3	2	0	4	4	3	0	0	3	4	3	0
Doc 4	2	1	1	0	1	1	3	0	3	0	0
Doc 5	3	3	4	1	3	0	3	4	4	2	1



	Passes in Final Third	Headed Shots	Interceptions	Turnover	Shot on Fastbreak	Flick-ons	Fouls
Match 1	26	1	3	0	0	3	4
Match 2	39	1	1	2	0	1	1
Match 3	12	1	3	2	1	4	2
Match 4	37	4	4	3	4	2	3
Match 5	31	4	2	0	3	0	3

# Latent Dirichlet Allocation (LDA)

- Generative Model



- (Sparse) Dirichlet Priors play an important role in the NLP implementation. Do they make sense for football?
- Bayesian Inference
  - Off-the-shelf MCMC sampling algorithms
  - We used the Sci-Kit Learn implementation

```
from sklearn.decomposition import LatentDirichletAllocation
```

# Notes on Implementation

- Pre-Processing:

- Traditionally *Tf-Idf*

$$tf(t, d) = (\text{raw count of term } t \text{ in document } d) / (\text{total number of terms in } d)$$

$$idf(t, D) = \log[(\text{total number of documents in the corpus } D) / (\text{documents in } D \text{ containing term } t)]$$

- Z-Score Scaler

- Parametric Choices:

- Number of topics and parameters of Dirichlet priors of “styles” (aka topics) and “style mixtures”
- Repeatability (throwback to MathSport 2017)

$$\left( \sum_{i \text{ in teams}} 2^{-\frac{\text{order}(v_i^1, v_i^2) + \text{order}(v_i^2, v_i^1)}{N-1}} \right) / N$$

- Empirical examination

## TEAM MODEL

Topic Label	Most Probable Features
<b>High Possession</b>	total pass, open play pass, accurate pass, leftside pass, rightside pass, touches, accurate back zone pass, backward pass, accurate forward zone pass, possession percentage, forward passes, successful long passes from own half into opposition's
<b>Dogged Defending</b>	Effective clearance, total clearance, effective head clearance, blocked cross, effective blocked cross, high claims, good high claims, lost corners, punches, outfielder block, attempts conceded out of box, possession won in defensive third, clean sheet, total launches
<b>Conceding Chances</b>	saves, diving saves, saved in box, attempts conceded in box, saved out of box, attempts conceded out of box, challenge lost, interceptions in box, outfielder block, error lead to goal, attempted tackle foul, lost corners, free kick given, yellow cards
<b>Crossing</b>	total cross, accurate cross, cross not from corner, corners into box. won corners, crosses behind 18 yards, crosses after 18 yards, penalty area entries, accurate crosses not from corners, shot off target, missed headed attempt, total headed attempts, missed attempt in box
<b>Fast Attacking</b>	attempted fast breaks, shot from fastbreak, total fast breaks, big chance created, one-on-one attempt, big chance scored, big chance missed, attempt from center of box, close miss, miis in box, shot off target, accurate through ball, on target scoring attempt
<b>Long Balls</b>	possession lost, possession lost control, total long balls, accurate launches, long pass from own half into opposition's, total launches, total flick ons, aerial won, aerial lost, accurate flick on, ball recovery, unsuccessful touch, duel won, duel lost, possession won middle third
<b>Shot Domination</b>	on target scoring attempt, accurate pull back, attempt on target right foot, total pull back, attempt on target in box, accurate through ball, total through ball, attempt saved in low centre, big chance created, attempt on target left foot, one-on-one attempt, attempt on target out of box, big chance scored, big chance missed, attempt from open play



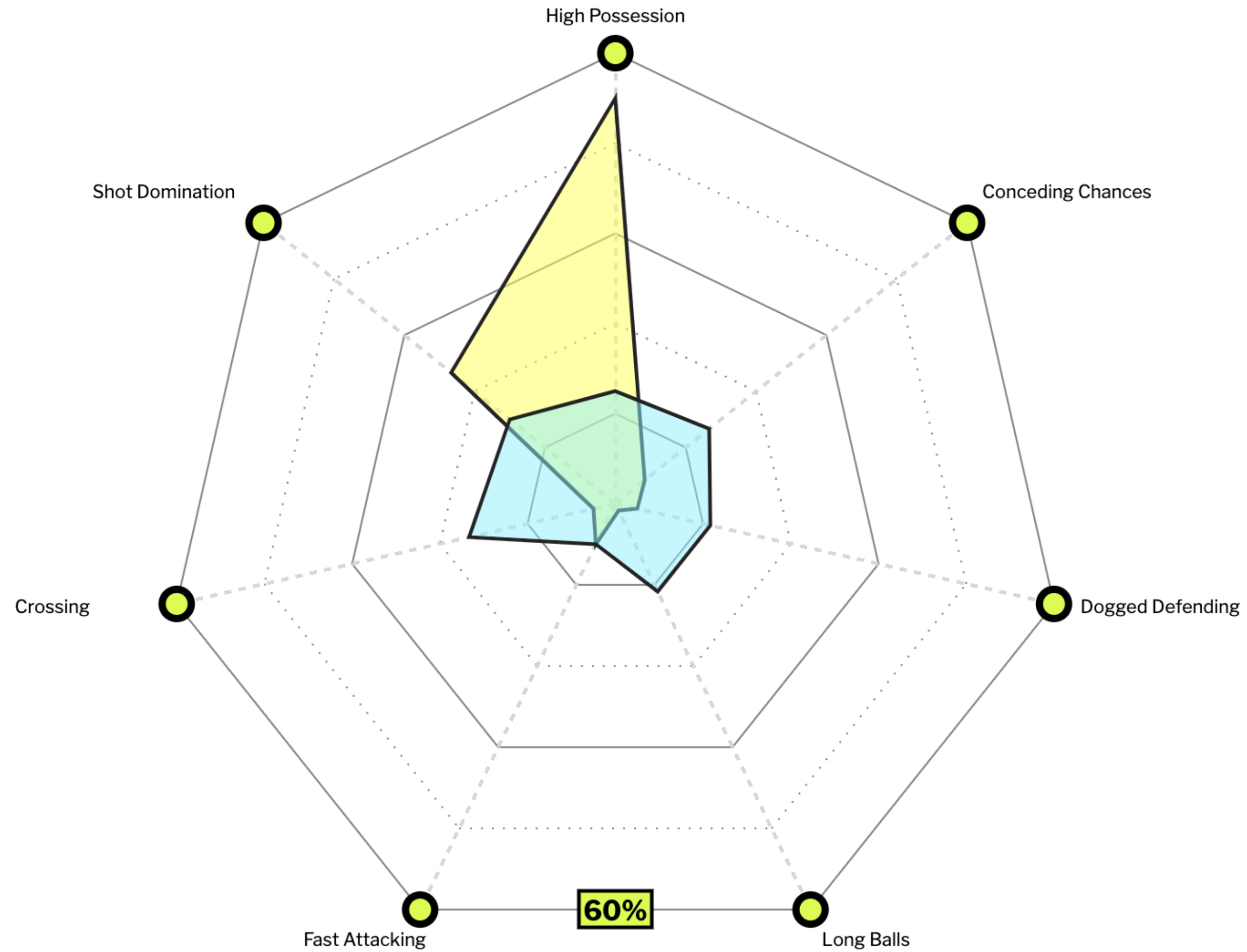


Barcelona



Atlético de Madrid

**How this works:** Our Persona model characterises teams **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.





Manchester City



Arsenal

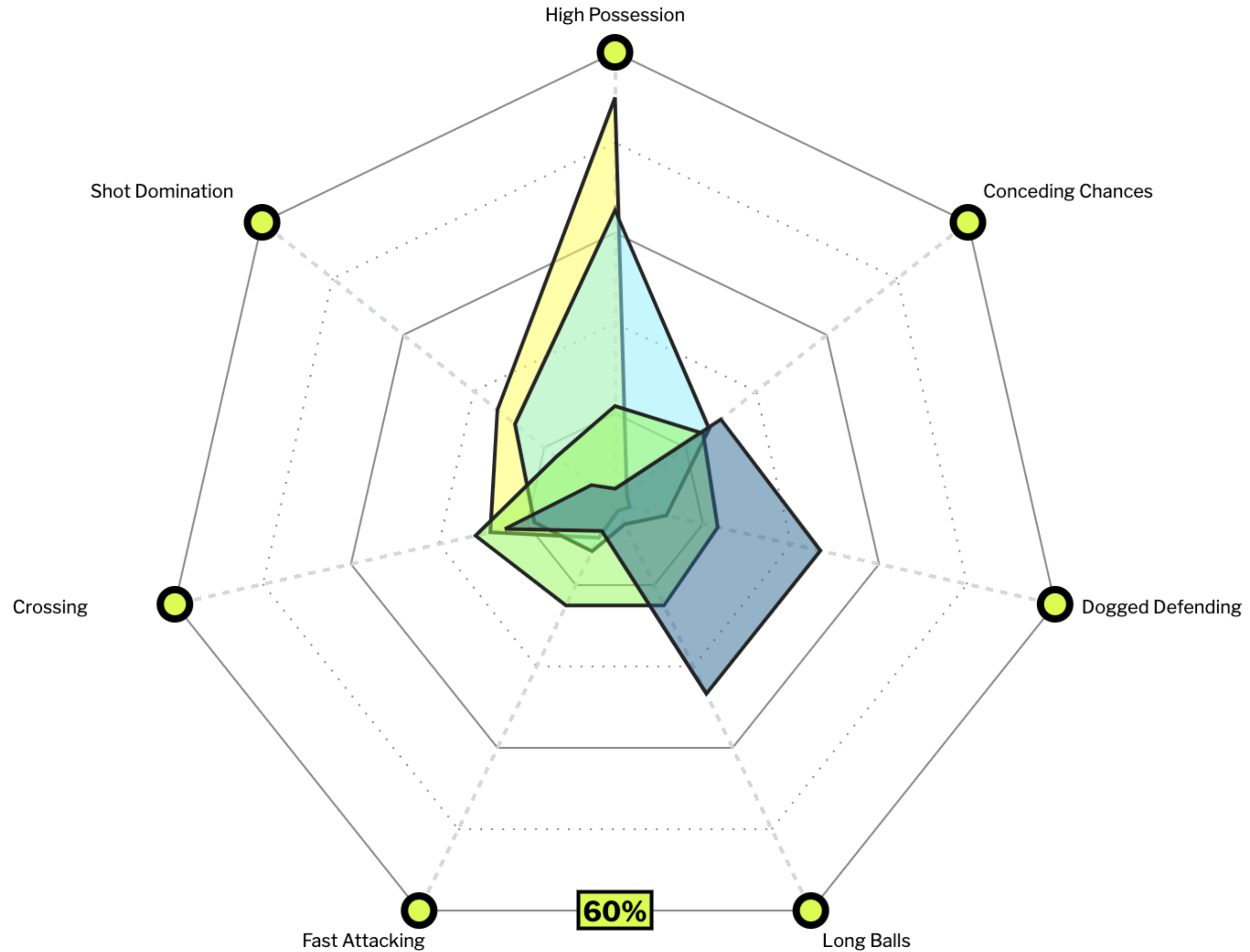


Wolverhampton Wanderers



Burnley

**How this works:** Our Persona model characterises teams **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



## DEFENDER MODEL

Topic Label	Most Probable Features
<b>Offensive Passing</b>	Passes left, rightside pass, accurate forward zone pass, touches, total forward zone pass, successful final third passes, open play pass, total pass, successful open play pass, accurate pass, final third entries, forward pass, backward pass, clean sheet, possession won middle third
<b>Build-Up Passing</b>	accurate back zone pass, total back zone pass, leftside pass, accurate pass, successful open play pass, total pass, open play pass, successful long passes from own half into opposition's, touches, forward pass, rightside pass, passes right, accurate forward zone pass, possession won defensive third, offside provoked, accurate long balls, clean sheet, head pass, ball recovery
<b>Active Defending</b>	Interceptions in box, interceptions, offside provoked, yellow card, total tackle, won tackle, possession won defensive third, outfielder block,, attempted tackle foul, fouls, challenge lost, duel won, ball recovery, was fouled, successful put through, head pass, possession won middle third, aerial lost
<b>Passive Defending</b>	effective clearance, effective head clearance, total clearance, outfielder block, offside provoked, aerial won, clean sheet, head pass, aerial lost, possession won defensive third, duel won, total launches, accurate launches, yellow card, total back zone pass, accurate back zone pass, long passes from own half into opposition's
<b>Blocking</b>	effective blocked cross, blocked cross, blocked pass, total tackle, won tackle, put through, possession won defensive third, duel won, attempted tackle foul
<b>Crossing</b>	crosses after 18 yards, total crosses not from corners, accurate cross not from corner, crosses behind 18 yards, total cross, penalty area entries, passes right, off target shot assist, possession lost, won corner, attempted assist from open play, total attempted assist, total final third passes, final third entries
<b>Long Balls</b>	Total chipped pass, accurate chipped pass, long pass from own half into opposition's, offside provoked, accurate long balls, successful long balls, forward pass, possession won defensive third, total launches, outfielder block, accurate launches, aerial won, total clearances

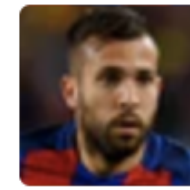
BETWEEN 1 AUGUST 2018 AND 1 JUNE 2019



**David Luiz**  
Chelsea



**Trent Alexander-Arnold**  
Liverpool

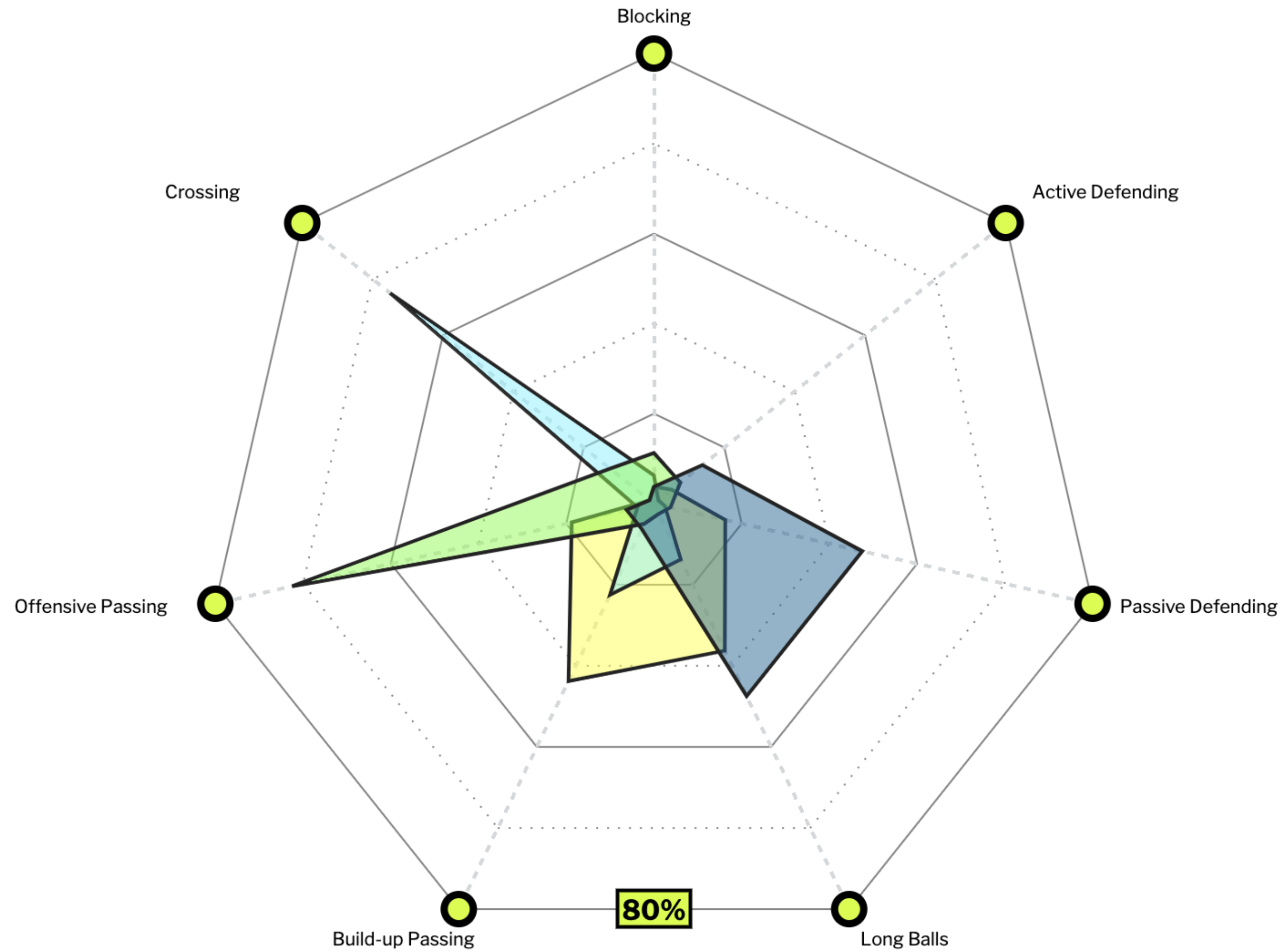


**Jordi Alba**  
Barcelona



**Ben Mee**  
Burnley

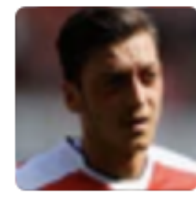
**How this works:** Our Persona model characterises players **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



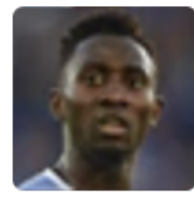
## MIDFIELDER MODEL

Topic Label	Most Probable Features
<b>Shooting</b>	Attempt to the centre from out of box, missed out of box attempt, blocked out of box attempt, total attempts with right foot, blocked scoring attempt, attempt open play, total scoring attempt, shot off target, total attempts left foot, on target attempts right foot, on target scoring attempts, missed attempt in box, won corners, touches in opposition box, possession won attacking third, successful final third passes
<b>Defending</b>	Total tackle, won tackle, attempted tackle foul, challenge lost, interception, interception won, yellow card, fouls, possession won middle third, ball recovery, possession won defensive third, outfielder block, duel won, successful put through, duel lost, clean sheet, blocked pass, chipped passes
<b>Build-Up Passing</b>	Total pass, accurate pass, open play pass, successful open play pass, accurate forward zone pass, touches, leftside pass, rightside pass, final third entries, accurate chipped pass, accurate back zone pass, total back zone pass, successful long passes from own half into opposition's, successful final third passes, forward pass, possession won middle third, accurate long balls
<b>Aerial Play</b>	Accurate flick ons, aerial won, head pass, total flick on, aerial lost, effective head clearance, head clearance, duel won, duel lost, was fouled, fouls, attempt from centre of box
<b>Dribbling</b>	Turnover, unsuccessful touch, overrun, total contest, won contest, blocked pass, put through, dispossessed, successful put through, duel lost, fouled final third, was fouled, possession lost, possession lost control, duel won, challenge lost, fouls, possession won attacking third, aerial lost
<b>Chance Creation</b>	Accurate layoffs, total layoffs, attempts assisted open play, on target attempts assisted, total attempts assisted, successful final third passes, total final third passes, backward pass, accurate forward zone pass, possession won attacking third, passes left, passes right, touches in opposition box, fouled final third, successful open play pass, penalty area entries
<b>Crossing</b>	Total crosses, accurate crosses, penalty area entries, accurate cross not from corner, crosses behind 18 yards, crosses after 18 yards, off target attempts assisted, total attempts assisted, won corner, possession lost control, on target attempt assisted, attempted assists open play, total forward zone passes, total final third passes, blocked pass, total contest

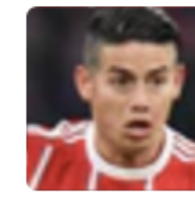
BETWEEN 1 AUGUST 2018 AND 1 JUNE 2019



**Mesut Özil**  
Arsenal



**Wilfred Ndidi**  
Leicester City

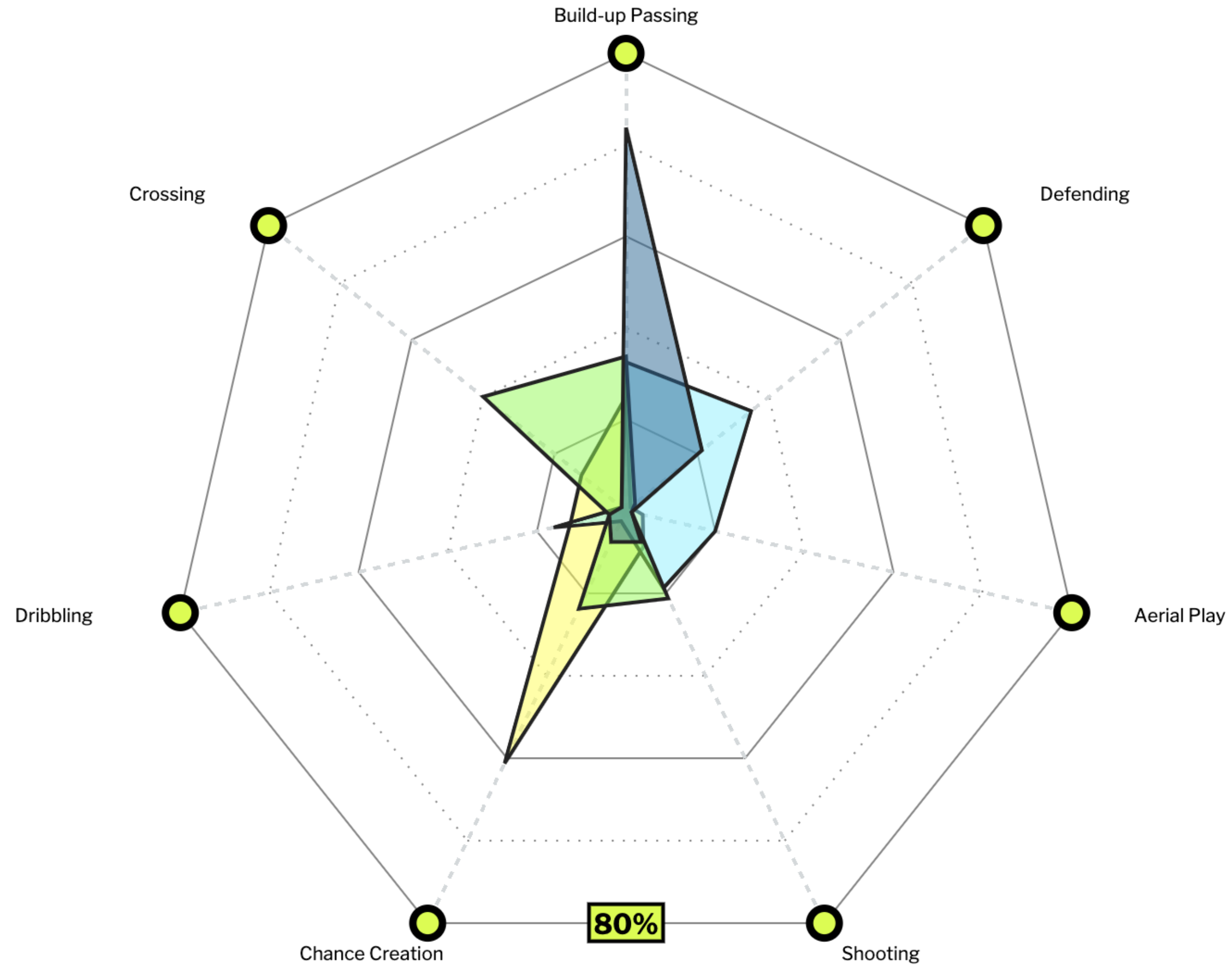


**James Rodríguez**  
Real Madrid



**Jorginho**  
Chelsea

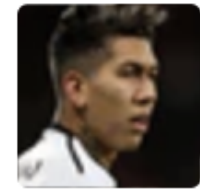
**How this works:** Our Persona model characterises players **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



## STRIKER MODEL

Topic Label	Most Probable Features
<b>Dribbling</b>	Unsuccessful touch, turnover, dispossessed, overrun, duel lost, fouled final third, total contest, was fouled, blocked pass, put through, won contest, fouls, successful put through, possession won attacking third, duel won, challenge lost
<b>Missed Shots</b>	Missed attempt in box, shot off target, attempted missed to the left, attempt missed to the right, missed headed attempt, total headed attempts, high missed attempt, total scoring attempts, big chance missed, attempt in box blocked, blocked scoring attempt, total offside
<b>Aerial Play</b>	Total flick on, accurate flick on, aerial lost, aerial won, head pass, duel lost, total offside, duel won, total layoffs, unsuccessful touch, turnover, accurate layoffs, fouls, dispossessed, total headed attempts, was fouled, possession lost control, effective head clearance
<b>Goalscoring</b>	Goals, goal inside box, goals open play, big chance scored, goal right foot, on target scoring attempt, attempt from centre of box, touches in opposition box, total scoring attempts, total offside, total attempts left foot, total headed attempts, total layoffs
<b>Chance Creation</b>	Accurate layoffs, total layoffs, total final third passes, successful final third passes, backward pass, accurate forward zone pass, total forward zone pass, attempt assisted open play, on target attempt assisted, total attempt assisted, passes left, rightside pass, won contest, touches in opposition box, off target attempt assisted, penalty area entries, touches
<b>On-Target Shots</b>	on target attempt in box, on target attempt right foot, attempt saved low centre, on target scoring attempt, on target attempt left foot, on target attempt out of box, attempt open play, total scoring attempt, attempt right foot, big chance missed, blocked scoring attempt, attempt blocked in box, touches in opposition box, attempt from centre of box, attempt from right of box, won corners
<b>Crossing</b>	Big chance created, accurate cross not from a corner, total attempted assist, accurate cross, crosses behind 18 yards, total crosses not from a corner, total cross, attempt assisted open play, off target attempt assisted, on target attempt assisted, penalty area entries, crosses after 18 yard

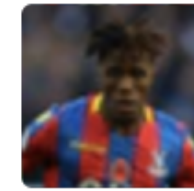
BETWEEN 1 AUGUST 2018 AND 1 JUNE 2019



**Roberto Firmino**  
Liverpool



**Chris Wood**  
Burnley

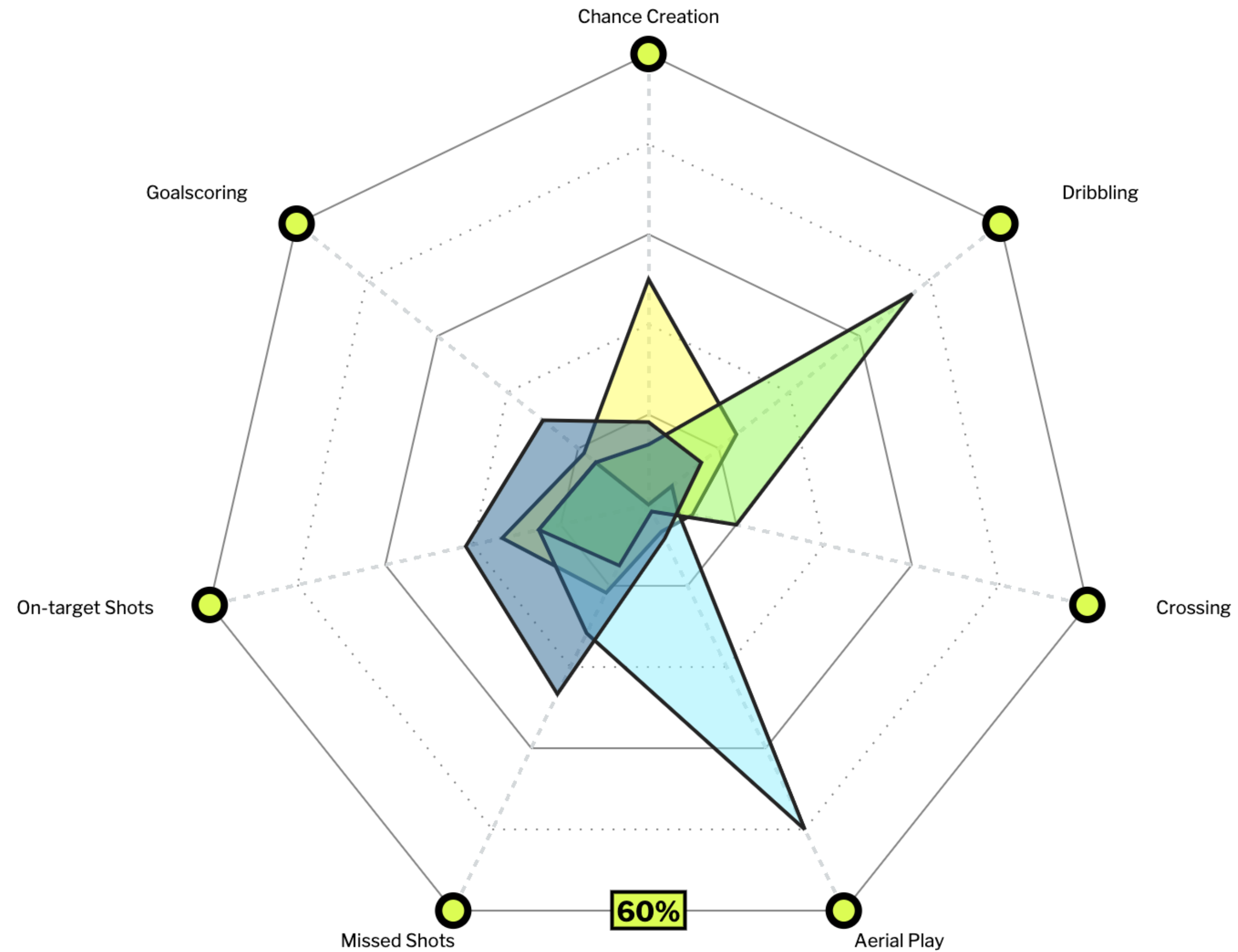


**Wilfried Zaha**  
Crystal Palace



**Sergio Agüero**  
Manchester City

**How this works:** Our Persona model characterises players **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.





# Application: Trend Tracking

TEAM COMPARISON: TIME PERIODS

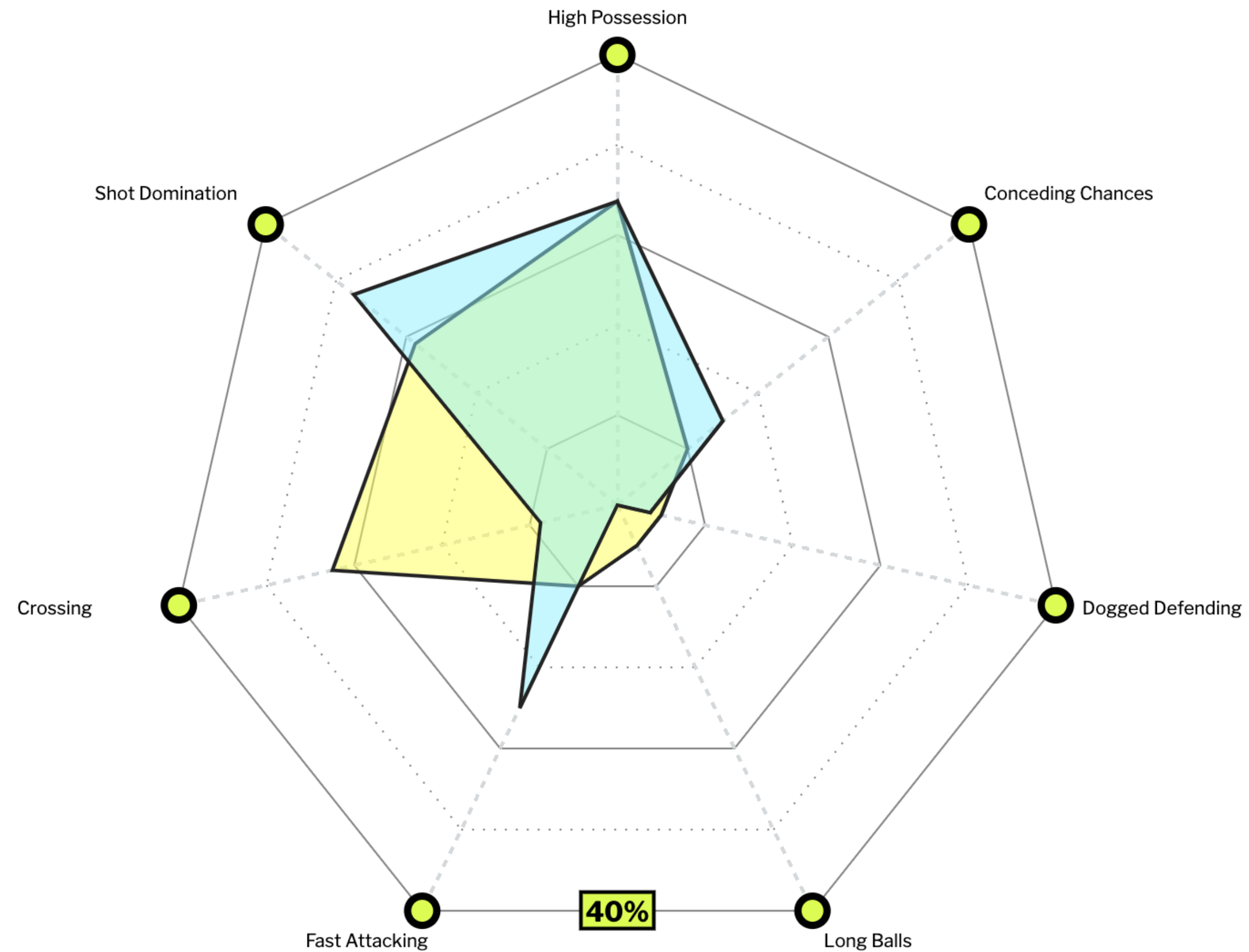


Lyon

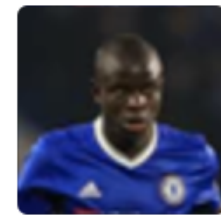
AUG 2018 - APR 2019

APR 2019 - MAY 2019

**How this works:** Our Persona model characterises teams **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



# Application: Changing Roles

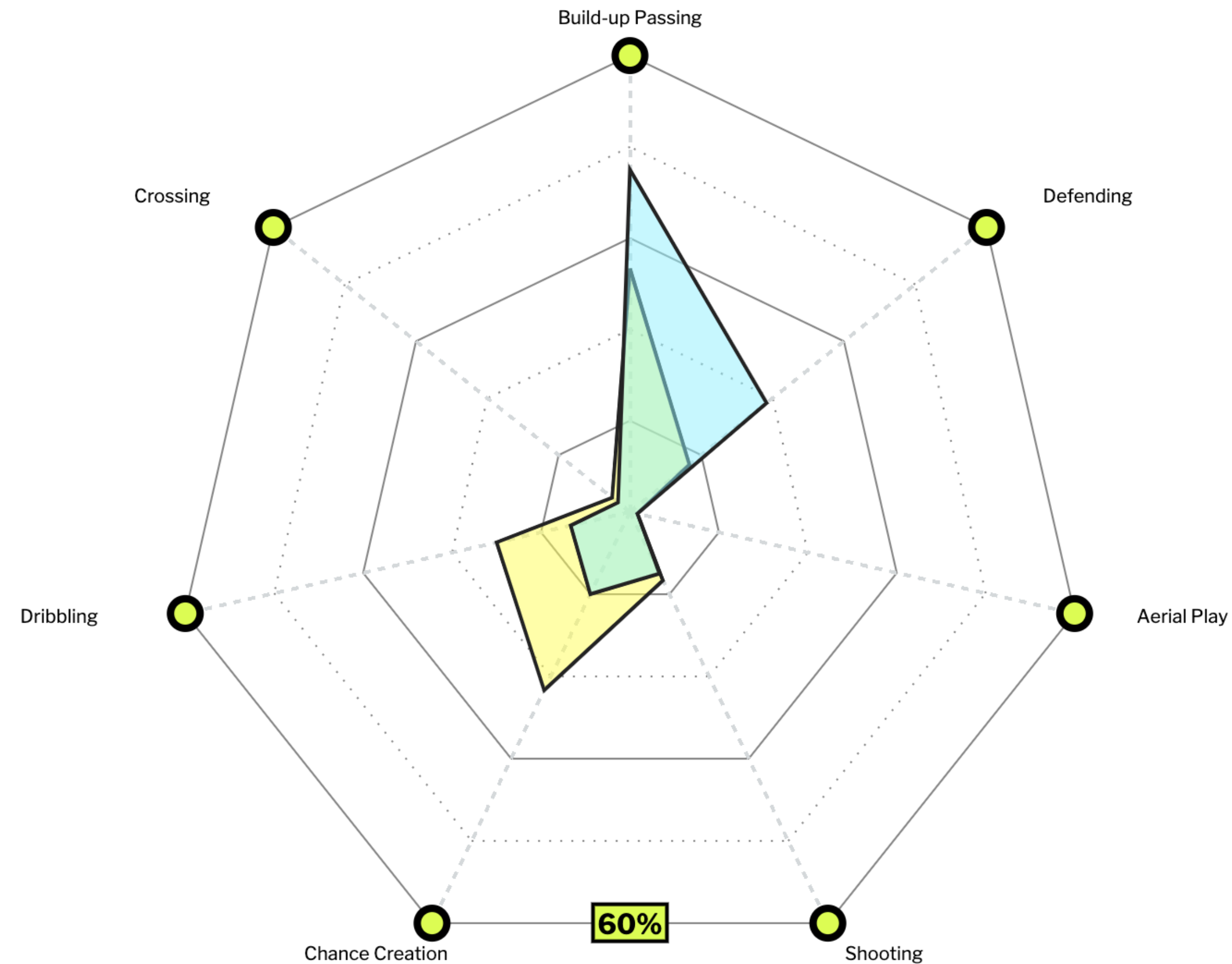


**N'Golo Kanté**  
Midfielder at Chelsea

AUG 2018 - MAY 2019

AUG 2017 - MAY 2018

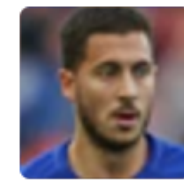
**How this works:** Our Persona model characterises players **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



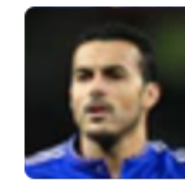
# Application: Similarity Recommender

BETWEEN 1 AUGUST 2018 AND 1 JUNE 2019

Player Name	Persona Similarity
Eden Hazard	100.0
Julian Brandt	97.4
Xherdan Shaqiri	97.0
Sam Larsson	96.1
Pedro	95.9
Lionel Messi	95.9
Paulo Dybala	95.8
Adem Ljajic	95.8
Cristian Tello	95.4
Neymar	94.9
David Neres	94.8
Ousmane Dembélé	94.5

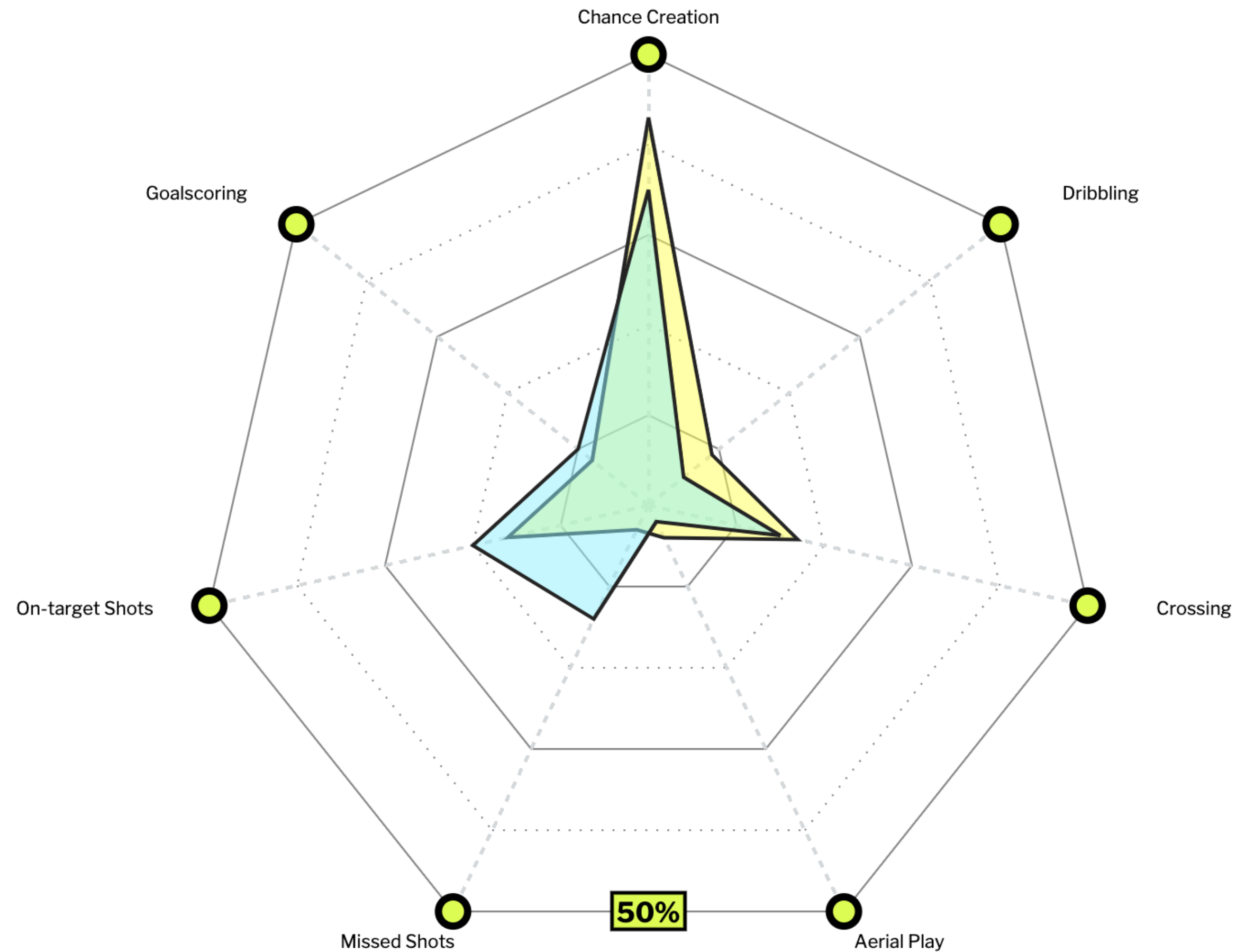


**Eden Hazard**  
Chelsea



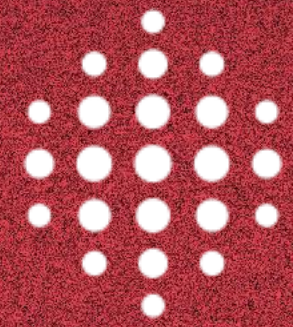
**Pedro**  
Chelsea

**How this works:** Our Persona model characterises players **stylistically** by grouping their output into event categories. They are assigned a % for each one respectively, which represents how much they perform that style **relative** to the others.



# Future Work

- **Implementation:**
  - Dirichlet Priors
  - Better Validation
- **Application:**
  - More Robust Metric in Encoding Space
  - Separate “context” from “style”. Benefit - there’s a direct relationship between team and player features



twenty3  
SPORTS INTELLIGENCE

Thank you

Questions?