Cultural evolution of football tactics: Strategic social learning in managers' choice of formation

Alex Mesoudi *

15 November, 2019

*Human Behaviour and Cultural Evolution Group, Department of Biosciences, College of Life and Environmental Sciences, University of Exeter, Penryn, Cornwall TR10 9FE, United Kingdom; a.mesoudi@exeter.ac.uk

Abstract

In order to solve complex problems or make difficult decisions, people must strategically combine personal information acquired directly from experience (individual learning) and social information copied from others (social learning). The game of football (soccer) provides a wealth of real world data with which to quantify managers' use of personal and social information in selecting team formations. I analyse a 5-year dataset of all games (n=9127, 2012-2017) in the five top European leagues (English Premier League, German Bundesliga, Spanish La Liga, French Ligue 1 and Italian Serie A) to quantify the extent to which a manager's initial match formation is guided by their personal past use or success with that formation, or other managers' use or success with that formation. I focus in particular on the use of the 4231 formation, which was the dominant formation at the start of this period but showed a gradual decline in use over time. I find that, as predicted, a manager's choice of whether to use 4231 or not is influenced by both their recent use of 4231 (personal information) and the use of 4231 in the entire population of managers in that division (social information). Contrary to expectations, managers appeared to rely more on personal than social information, although this estimate was highly variable across managers and divisions. Serie A managers, in particular, showed a much stronger use of personal information, likely due to the rarity of 4231 in that division. Finally, there did not appear to be an adaptive tradeoff between social and personal information use, with the relative reliance on each failing to predict managerial success.

Introduction

When solving problems or making decisions, people use a combination of personal information acquired directly from the environment (individual learning), and social information acquired by copying others (social learning)[1–7]. The strategic combination of individual and social learning is adaptive when decisions or problems are challenging, such as when environments change over time such that social information may become outdated[2,3,7], or when solutions are causally opaque or multidimensional, such that they cannot be acquired by individual learning alone and require the social learning of accumulated past solutions[1,2]. People show this strategic mix of individual and social learning in the lab[8–13] and the real world[14,15] (although sometimes imperfectly[16]). When combined appropriately, individual and social learning can generate cumulative cultural evolution, where innovations generated via individual learning are preserved and accumulated over generations via social learning[17].

Beheim, Thigpen & McElreath[14] provided an innovative demonstration of the strategic use of social and individual learning in the real world. They analysed decades of professional matches of the board game Go to understand the spread of an opening move, the "Fourfour". This move appeared in 1968 and then increased rapidly in frequency. Beheim et al. showed that Go players' use of Fourfour is predicted by both personal information - the past use and win rate of Fourfour by that player - and social information - the past use and win rate of Fourfour in the entire population of Go players. They also showed considerable between-player variation, with some players using predominantly social information (e.g. Lee Sedol), and others using mostly personal information (e.g. Takemiya Masaki, the originator of Fourfour).

Here I apply the methods and approach of Beheim et al. to another competitive real world sport, football (soccer). Football is enjoyed by millions of people worldwide, and European leagues alone have a revenue of almost €30 billion[18]. Football has been subject to historical analyses of tactics[19] and increasingly, by providing a wealth of fine-grained quantitative data, statistical analyses[20].

The equivalent in a football match to a Go player's opening move is a manager's starting formation. This describes how the ten outfield players are initially organised on the pitch. Formations are typically defined by three or four numbers specifying the number of players in each segment of the pitch. For example, 442 comprises four defenders, four midfielders and two attackers. While formations may change during matches in response to player substitutions or other in-game events, all managers select one of a finite and, in practice, relatively small set of starting formations. Formations are a key component of overall tactics. For example, 541 is more defensive than 343. Historically, formations and tactics have gradually evolved over time, with noticeable long-term trends and short-term cycles[19].

The key question addressed here is the extent to which managers use personal and social information to decide on their starting formation. This is a challenging decision, as defined above. The success of a formation depends partly on what formation the opposing manager plays, making payoffs of the same formation temporally variable and frequency dependent. Various other factors, from squad strength to luck, determine match outcomes in addition to formation, making the true contribution of the latter difficult to determine. And in the high stakes of football management (the median tenure of English Premier League managers as of August 2019 was 1 year, 158 days), there are limited opportunities to directly trial formations, especially if those trials are unsuccessful.

To maintain tractability and comparability to Beheim et al., I examine a manager's choice of whether to play the 4231 formation or not. 4231 was the dominant formation during the period of study (Figure 1): in the top five European domestic leagues from 2012-2017, it was used 37% of the time, more than double the next most common formations (18% for 433 and 14% for 442). However, 4231 also showed a clear decline in frequency during this period, from 47% in the 2012/13 season to 28% in the 2016/17 season. This decline was more extreme in some leagues than others; for example, the Spanish La Liga saw a decline in 4231 use from 78% to 37%.

Here I use a 5-year dataset of all games (n=9127, 2012-2017) in the five top European leagues (English Premier League, German Bundesliga, Spanish La Liga, French Ligue 1 and Italian Serie A) to test the following hypotheses, derived from the above theory and the results of Beheim et al.[14]. All hypotheses and analyses were preregistered before running any analyses on the original data (https://osf.io/er4dx), and all data and code are available at https://github.com/amesoudi/football.

- H1: A manager's choice of whether to play 4231 is determined by a combination of personal

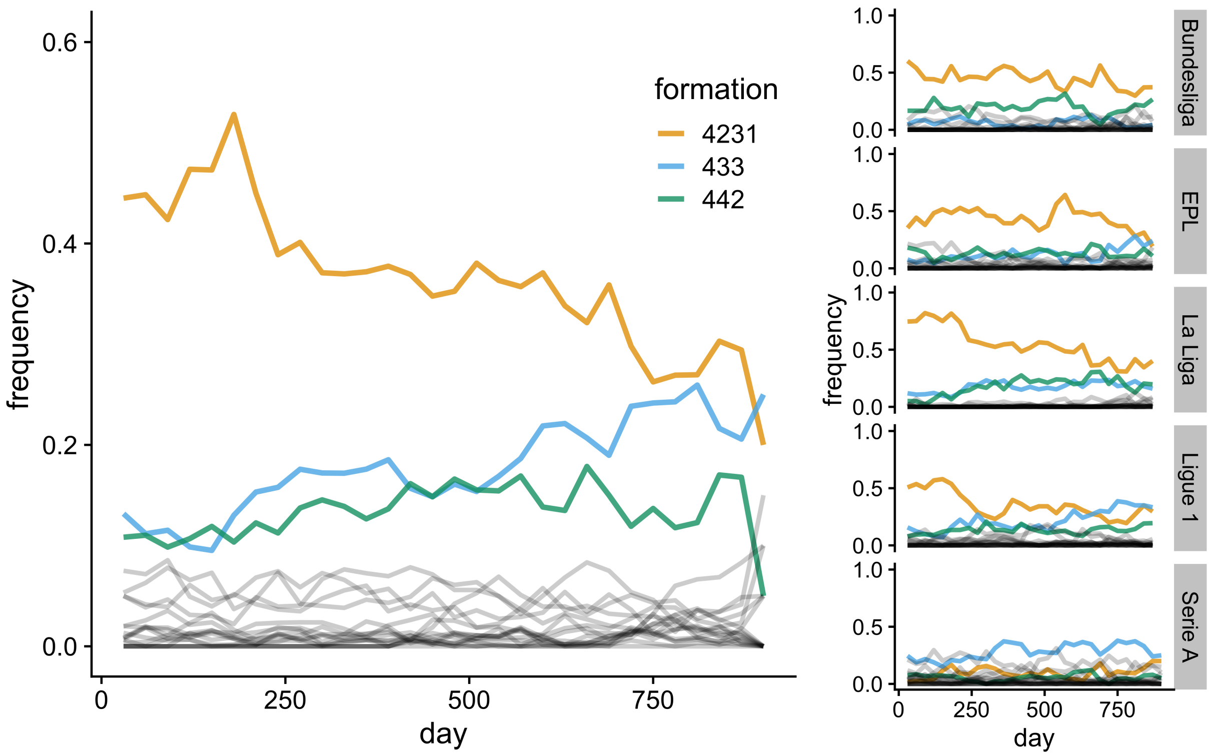

Figure 1: Frequencies of initial formations across all leagues (large image) and in the five separate leagues (right panels). The three most common formations are shown: 4231 (orange), 433 (blue) and 442 (green). Other less common formations are shown in grey. Frequencies are calculated as the proportion of all matches in consecutive 30-day bins that started with that formation. 'Days' are consecutive match days across five seasons from 2012-2017, omitting days on which no matches were played.

and social information.

- H2: On average, there is greater reliance on social than personal information, as found by Beheim et al. for Go players.
- H3: There is more variation between managers in both personal and social information use compared to randomised data.
- H4: There is an n-shaped relationship between the ratio of population:personal information use and a manager's success, indicating that an overreliance on either form of information is less effective than strategically combining the two.

Methods

Data

The original dataset was downloaded from the website Kaggle, originally compiled by Jemilu Mohammed from various online sources including whoscored.com, dated 6th July 2017 (version 3) and with licence CC0: Public Domain. There are 9127 games in the dataset, which gives 18254 starting formations (two per game, one for each team). The downloaded dataset is available as Supplementary Information.

Data was preprocessed to correct inconsistent spelling of manager names (e.g. Arsène Wenger and Arsene Wenger were merged, as were Gus Poyet and Gustavo Poyet), add one missing formation and one missing manager, add season indicators using official season start and end dates, and create predictor variables.

Predictors

Following Beheim et al., predictors were created using a moving time window of X days previous to the formation choice in question. That is, for each formation choice, predictors were calculated from all games played in the previous X days, not including the day on which that formation was played. In the analyses reported below $X = 30$, and analyses are repeated in the Supplementary

Information using $X = 20$, $X = 40$ and $X = 60$ (these choices were preregistered, and generated qualitatively identical results to those presented below for $X = 30$). The X-day window was reset at the start of each season, so games played in the first X days of each season were not included in the analyses. The Bundesliga, unlike the other leagues, has a mid-season break of more than 30 days; this was reduced to 10 days in each season so that all X-day windows yielded prior game data.

Personal predictors were (a) personal use of 4231: the proportion of games in the X-day window played by that manager in that division and season in which 4231 was chosen, out of all games played by that manager in that division and season in the X-day window, centred on 0.5; (b) personal 4231 win rate: the proportion of games played with 4231 by that manager in that division and season in the X-day window that were won, centred on the equivalent win rate for games played with a non-4231 formation; (c) the interaction between personal 4231 use and personal 4231 win rate; and (d) the interaction between personal 4231 use and the managers' overall win rate with any formation, with the latter centred on the overall win rate of all managers in that division and season.

Population predictors were (e) population 4231 use: the proportion of games in the X-day window played in that division and season in which 4231 was chosen, out of all games played in that division and season in the X-day window, centred on 0.5; (f) population 4231 win rate: the proportion of games played with 4231 in that division and season in the X-day window that were won, centred on the equivalent win rate for games played with a non-4231 formation; (g) the interaction between population 4231 use and population 4231 win rate; and (h) the interaction between population 4231 use and the managers' overall win rate with any formation, with the latter centred on the overall win rate of all managers in that division and season.

Additional predictors were an indicator variable denoting whether the formation was used home or away, and a measure of team strength which was the proportion of games won by that team in that entire season, centred on the mean win rate of all teams in that division in that season.

Analyses

Bayesian multi-level regression models were run using the rethinking package and book[21,22]. All models contained varying effects for manager and division. The null model contained only the home/away and team strength predictors. The personal model contained home/away, team strength and the four personal information predictors. The population model contained home/away, team strength and the four population information predictors. The full model contained home/away, team strength and all eight personal and population variables. In addition to varying intercepts for manager and division, the personal model contained varying slopes for personal information use and win rate, the population model contained varying slopes for population information use and win rate, and the full model had both sets of varying slopes. Full model specifications can be found in the Supplementary Information.

Predictions

Hypotheses H1-H4 specified in the Introduction were tested statistically via the following predictions:

H1 predictions: The full regression model with personal (individual) and population (social) predictors has better fit to the data than the personal-only model, the population-only model, and the null model with neither personal nor population predictors. Fit is indicated by model comparison using WAIC. Additionally, in the full model, there are effects of (a) personal 4231 use, (b) personal 4231 win rate, (c) population 4231 use and (d) population 4231 win rate, and interactions between (e) personal 4231 use and win, and between (f) population 4231 use and win rates. Effects are indicated by the parameter estimates' 89% CI (specifically, 89% percentile interval, see[22]) not including zero in the full model.

H2 prediction: The ratio of population:personal use, calculated by dividing the estimate for population use in the full model by the estimate for personal use in the full model, is reliably greater than one, as indicated by 89% CIs not overlapping one.

H3 predictions: The standard deviation of the varying effects for managers' (a) personal 4231 use and (b) population 4231 use in the best-supported regression model will be larger than the equiv-

Table 1: Model comparison to test hypothesis H1.

	WAIC	pWAIC	dWAIC	weight	SE	dSE
Full model	12273.11	338.30	0.00	1	144.03	NA
Personal model	12346.35	298.20	73.24	0	144.19	15.16
Population model	13547.72	356.86	1274.61	0	137.23	88.05
Null model	14331.91	251.01	2058.79	0	135.11	101.69

alent standard deviations in a model generated with dummy data that has randomised formation and win rates across managers.

H4 prediction: In a regression model with manager as unit of analysis which contains the managers' population:personal information use ratio and their personal win rate relative to other managers in that division and season, the win rate is reliably predicted by the square of the population:personal use ratio (i.e. a negative coefficient in a quadratic polynomial).

Results

H1 predictions: combination of population and personal information use

As predicted, the full model containing both personal and population predictors was best supported, containing all of the model weight compared to the personal, population and null models (Table 1). However, the WAIC of the personal model came much closer to the full model WAIC than the population or null models.

Table 2 shows the parameter estimates for the full model. As predicted, there are effects of personal 4231 use, personal 4231 win rate and an interaction between these two predictors. There is also an effect of population 4231 use. However, there were no reliable effects of population 4231 win rate, nor interactions between personal 4231 use and personal overall win rate, nor interactions between population 4231 use and either population 4231 win rate or personal overall win rate.

Figure 2 shows how past personal 4231 use (Figure 2A) and population 4231 use (Figure 2B) increase the probability of 4231 being chosen. The interaction between personal 4231 use and personal 4231 win rate revealed in Table 2 can be seen in Figure 2A: for managers who have seldom used 4231 in recent games, a higher win rate with 4231 increases their likelihood of choosing 4231,

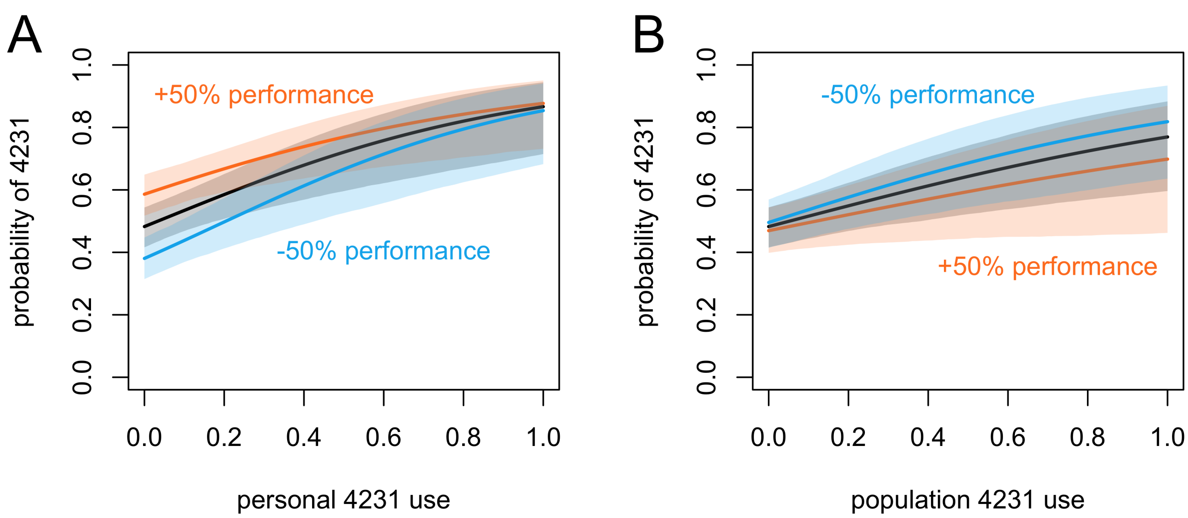

Figure 2: (A) The predicted probability of using 4231 as a function of personal 4231 use, assuming that the 4231 win rate is the same as the non-4231 rate (black line, grey shading showing 89% CI), assuming that the personal 4231 win rate is 50% higher than the non-4231 win rate (orange line and shading), and assuming that the personal 4231 win rate is 50% lower than the non-4231 win rate (blue line and shading). (B) The equivalent predictions for population 4231 use, and +50% or -50% population 4231 win rates relative to non-4231 population win rates.

Table 2: Parameter estimates for the full model. Home/away is an indicator trait with separate estimates for formations used home and away. Varying effects show the standard deviations of the varying intercepts and slopes. See SI for full model specification and priors.

	mean	sd	5.5%	94.5%
Fixed effects:				
Home/away: Home	-0.01	0.17	-0.27	0.24
Home/away: Away	-0.14	0.17	-0.41	0.11
Team strength	0.03	0.27	-0.39	0.46
Personal 4231 use	2.08	0.63	0.96	2.92
Personal 4231 win rate	0.84	0.13	0.63	1.06
Personal 4231 use * personal 4231 win rate	-0.64	0.21	-0.98	-0.30
Personal 4231 use * personal win rate	0.11	0.35	-0.46	0.66
Population 4231 use	1.34	0.48	0.55	2.03
Population 4231 win rate	-0.11	0.20	-0.43	0.22
Population 4231 use * population 4231 win rate	-0.62	0.67	-1.69	0.45
Population 4231 use * personal win rate	-1.07	0.51	-1.87	-0.24
Varying effects:				
Manager	0.71	0.07	0.60	0.83
Manager * personal 4231 use	1.26	0.14	1.04	1.50
Manager * population 4231 win rate	2.00	0.42	1.35	2.67
Division	0.27	0.19	0.05	0.62
Division * personal 4231 use	1.25	0.59	0.57	2.32
Division * population 4231 win rate	0.69	0.53	0.06	1.66

and a lower win rate decreases that likelihood. Also consistent with Table 2, Figure 2B reveals no reliable interaction between population 4231 use and population 4231 win rate, with the higher and lower performance lines falling within the average performance CI shading.

H2 predictions: ratio of population to personal 4231 use

The ratio of population:personal use, as calculated using the full model (see Table 2), had a mean of 0.68 (89% CI[0.24, 1.41]). Contrary to hypothesis H2, this suggests that personal information was more influential than population (i.e. social) information. As indicated by the wide confidence intervals, however, this estimate was highly uncertain, and there was a lot of variation in this ratio across managers and divisions.

Table 3: Tests of the differences between varying effects from the real data and varying effects from randomised data, to test hypothesis H3. Values shown are real minus randomised standard deviations.

	mean	sd	5.5%	94.5%
Manager * personal 4231 use	1.07	0.19	0.78	1.37
Manager * population 4231 use	1.78	0.46	1.04	2.48
Division * personal 4231 use	1.16	0.59	0.45	2.24
Division * population 4231 use	0.46	0.58	-0.27	1.47

Table 4: Model estimates for the quadratic regression model with manager as unit of analysis, to test hypothesis H4. Parameter a is the intercept, b1 is the linear coefficient and b2 the quadratic coefficient. Win rate is modelled as normally distributed with standard deviation sigma. See SI for priors.

	mean	sd	5.5%	94.5%
a	0.03	0.14	-0.19	0.26
b1	-0.16	0.18	-0.44	0.11
b2	0.10	0.07	-0.01	0.21
sigma	1.00	0.04	0.94	1.06

H3 predictions: variation across managers and divisions

Table 3 shows that, as predicted, there was more variation across managers in the effects of both personal 4231 use and population 4231 use compared to randomised data. Also as predicted, there was more variation across divisions in the effect of personal 4231 use compared to randomised data, but not in the effect of population 4231 use.

H4 predictions: population to personal use ratio and win rate

Contrary to hypothesis H4, there was no n-shaped quadratic relationship between manager win rate and population:personal information use ratio (Table 4 and Figure 3). There was also no reliable linear relationship: as shown in Figure 3, for most use ratios the relationship with win rate is flat. Managers with very high ratios, indicating more reliance on population 4231 use than personal 4231 use, had higher win rates, but the shaded 89% CIs always included zero, and this increase is likely unduly influenced by the right-most outlier.

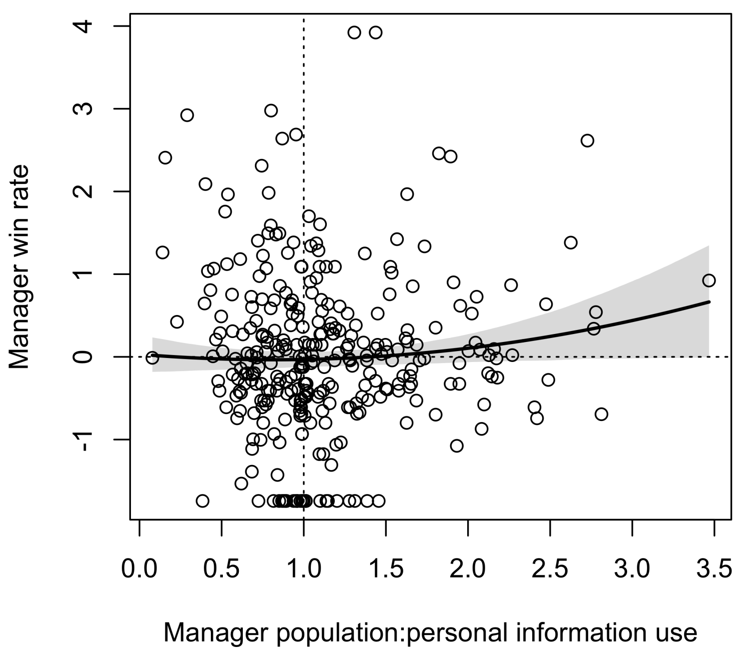

Figure 3: Relationship between each manager's win rate relative to the average manager's win rate and each manager's population:personal information use ratio as generated from the full model. Dotted lines indicated the average win rate and equal ratio. The thick line shows the predicted mean win rate at each value of the ratio, with shaded 89% CIs.

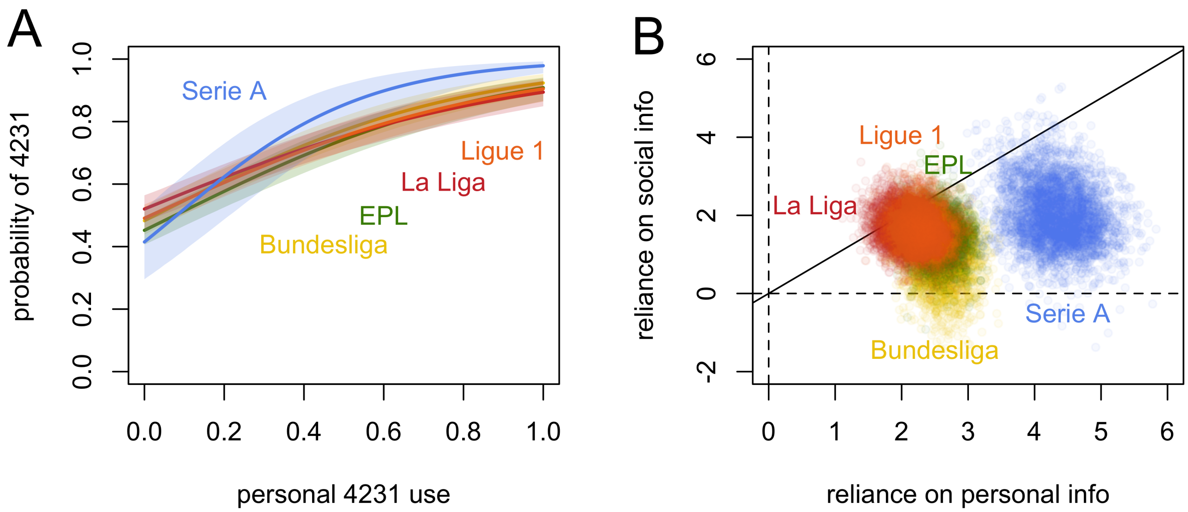

Figure 4: (A) Effect of personal 4231 use on probability of choosing 4231 broken down by division. (B) Joint posterior densities of the relative reliance on personal and social information, for the five divisions. The solid black diagonal indicates equal personal and population influence. EPL = English Premier League.

Exploratory analysis: between division effects

Figure 4 shows the variation across the five divisions in the effect of personal 4231 use. Four of the divisions are almost identical. The Italian Serie A, however, shows a much stronger effect of personal information use. This is likely because of the low overall frequency of 4231 in this division, as shown in the Serie A inset in Figure 1. The majority of managers in Serie A never or seldom used 4231 during this period; out of 67 managers who managed in Serie A, only three used 4231 in more than 50% of their games, only ten used it in more than 25% of their games, and 35 never used it. The small number of managers who used 4231 in a majority of their games would have disproportionately influenced the model's predicted probability of subsequently picking 4231 at high values of personal 4231 use, as seen in Figure 4A.

Exploratory analysis: between manager effects

Figure 5A shows variation across five successful managers who played over 100 games in our study period: Antonio Conte (78% win rate, 101 games), Josep Guardiola (74% win rate, 124 games),

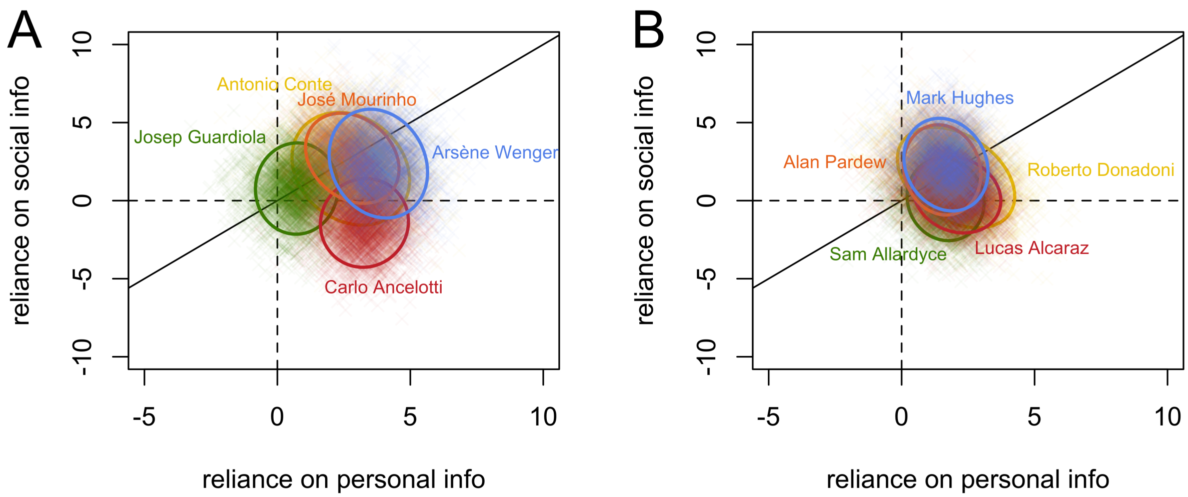

Figure 5: Joint posterior densities of the relative reliance on personal and social information, for (A) five managers with high win rates and (B) five managers with low win rates, all of whom have managed more than 100 games in the period of study. Ellipses indicate the 80% confidence region for each manager. The solid black diagonals indicate equal personal and population influence.

Carlo Ancelotti (73%, 131 games), José Mourinho (59% win rate, 149 games) and Arsene Wenger (59% win rate, 170 games).

Figure 5B shows variation across five relatively unsuccessful managers who played over 100 games in our study period: Roberto Donadoni (30% win rate, 161 games), Sam Allardyce (32% win rate, 151 games), Alan Pardew (34%, 147 games), Lucas Alcaraz (27% win rate, 107 games) and Mark Hughes (35% win rate, 144 games).

While exploratory, we can see in these figures that successful managers seem to be more different to one another in information use strategies than unsuccessful managers. Carlo Ancelotti has less reliance on social information than the other successful managers, while Josep Guardiola has relatively less reliance on personal information. The unsuccessful managers show substantial overlap with one another over a smaller combined area than the successful managers. Whether this pattern is robust, and the reasons for it, are worthy of further study. Perhaps there are more ways to be successful than there are to be unsuccessful in football management.

Discussion

Complex decisions often require the strategic combination of personal information acquired via individual learning and population-wide information acquired via social learning, each of which has distinct advantages and disadvantages. Beheim et al.[14] analysed decades of games of Go to show that professional Go players combine personal and social information when deciding on opening moves, and these individual-level strategic decisions generated long-term evolutionary dynamics. Here, I applied the same methodological approach to the game of football, where the equivalent to an opening Go move is a manager's choice of starting formation. Consequently, I examined personal and social influences on a manager's choice of whether to use the most popular 4231 formation or not.

Over five seasons from 2012-2017 across the five top European leagues, it is indeed the case (supporting Hypothesis H1) that a manager's choice of whether to play 4231 is on average determined by both their own recent use of 4231 (personal information) and the frequency within which 4231 is recently used in the entire population of managers from the same league (social information), as well as the manager's personal win rate with 4231.

Contrary to the more specific prediction (Hypothesis H2) that managers should rely more on social than personal information, given the difficulty of personally trialling different formations in the high stakes world of football management and previous findings to that effect by Beheim et al.[14], there was if anything more reliance on personal information. This is puzzling not only for the aforementioned reasons (the difficulty of individual learning should favour reliance on social learning, plus the previous findings of Beheim et al.), but also the fact that the population provides much more information overall in the same time period: for the 30-day time window used here, a manager using population-wide information can draw on a mean of 76 games played across the entire division, while personal information only provides data from a mean of 3.6 games. The preference for personal information may be evidence for an egocentric bias, with managers weighting their own experience higher than others' experience. Previous lab experiments have found similar over-reliance on individual learning at the expense of social learning[9,11,13,16,23,24].

On the other hand, such estimates of the relative reliance on personal and population information were very uncertain, with confidence intervals so wide as to be consistent with a reliance on either form. This is due to the extensive variation in information use strategy across both managers and divisions, much more than would be expected if decisions were random (Hypothesis H3).

However, this variation does not seem to exhibit an adaptive tradeoff (Hypothesis H4): managers with different ratios of population:personal information use did not vary in their success, contrary to the expectation that an overreliance on either type of information should be detrimental. Perhaps in team sports like football, starting formations do not reliably translate into success in the way that opening moves in Go do, given the many other factors that determine success and the possibility of changing formations during a game.

Exploratory analyses showed that one league, Serie A, showed a much stronger effect of personal information than the other leagues. This is likely because of the low overall use of 4231 in this league, with very few managers using this formation; this small number of managers drove the effect, given that if a manager used 4231 previously, they were highly likely to be one of the few managers to use it in the future. This illustrates two points: first, the importance of including league (or any other grouping variable) as a varying effect in the analysis, to account for unusual patterns such as this, and second, the influence of overall trait frequency on the learning strategies that are employed. Rare traits may be driven more by personal experience, when a manager is unable to draw on the experience of others.

Future analyses might apply more generative models to data such as this, for example reinforcement, memory decay or Bayesian updating models to more directly model how managers might be updating their beliefs in response to constantly changing personal and social information. This may necessitate different implementations of the time window. Here this was a fixed X-day time window preceding the formation choice. More advanced models might include all previous games in a season/division, weighted by recency, or even social network ties amongst managers.

References

1. Boyd, R. & Richerson, P. J. *Culture and the evolutionary process.* (Univ. Chicago Press, 1985).

2. Boyd, R. & Richerson, P. J. Why does culture increase human adaptability? *Ethology and Sociobiology* **16,** 125–143 (1995).

3. Enquist, M., Eriksson, K. & Ghirlanda, S. Critical social learning: A solution to Rogers' paradox of nonadaptive culture. *American Anthropologist* **109,** 727–734 (2007).

4. Kendal, R. L. *et al.* Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences* (2018). doi:10.1016/j.tics.2018.04.003

5. Laland, K. N. Social learning strategies. *Learning & Behavior* **32,** 4–14 (2004).

6. Perreault, C., Moya, C. & Boyd, R. A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior* **33,** 449–459 (2012).

7. Rogers, A. Does biology constrain culture? *American Anthropologist* **90,** 819–831 (1988).

8. Kameda, T. & Nakanishi, D. Does social/cultural learning increase human adaptability? Rogers' question revisited. *Evolution and Human Behavior* **24,** 242–260 (2003).

9. McElreath, R. *et al.* Applying evolutionary models to the laboratory study of social learning. *Evolution and Human Behavior* **26,** 483–508 (2005).

10. Mesoudi, A. An experimental simulation of the 'copy-successful-individuals' cultural learning strategy: Adaptive landscapes, producer-scrounger dynamics and informational access costs. *Evolution and Human Behavior* **29,** 350–363 (2008).

11. Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W. & Laland, K. N. The evolutionary basis of human social learning. *Proceedings of the Royal Society B* **279,** 653–662 (2011).

12. Toelch, U. *et al.* Decreased environmental variability induces a bias for social information use in humans. *Evolution and Human Behavior* **30,** 32–40 (2009).

13. Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J. & Reader, S. M. Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B* **281,** 20132864 (2014).

14. Beheim, B. A., Thigpen, C. & McElreath, R. Strategic social learning and the population dynamics of human behavior: The game of Go. *Evolution and Human Behavior* **35,** 351–357 (2014).

15. Miu, E., Gulley, N., Laland, K. N. & Rendell, L. Innovation and cumulative culture through tweaks and leaps in online programming contests. *Nature Communications* **9,** 2321 (2018).

16. Mesoudi, A. An experimental comparison of human social learning strategies: Payoff-biased social learning is adaptive but underused. *Evolution and Human Behavior* **32,** 334–342 (2011).

17. Mesoudi, A. & Thornton, A. What is cumulative cultural evolution? *Proceedings of the Royal Society B* **285,** 20180712 (2018).

18. Barnard, M., Boor, S., Winn, C., Wood, C. & Wray, I. *World in motion: Annual review of football finance 2019.* (Sports Business Group, Deloitte, 2019).

19. Wilson, J. *Inverting the pyramid: The history of football tactics.* (Bold Type Text, 2013).

20. Tamura, K. & Masuda, N. Win-stay lose-shift strategy in formation changes in football. *EPJ Data Science* **4,** 1–19 (2015).

21. McElreath, R. Rethinking: Statistical Rethinking Book Package version 1.88. (2019).

22. McElreath, R. *Statistical rethinking: A Bayesian course with examples in r and stan.* (CRC Press, 2016).

23. Efferson, C., Lalive, R., Richerson, P. J., McElreath, R. & Lubell, M. Conformists and mavericks: The empirics of frequency-dependent cultural transmission. *Evolution and Human Behavior* **29,** 56–64 (2008).

24. Weizsacker, G. Do we follow others when we should? A simple test of rational expectations. *American Economic Review* **100,** 2340–2360 (2010).