Information Transfer and Home Field **Advantage**

INTRODUCTION

One of the most intriguing elements in professional sports is the home field advantage. Home field advantage often produces unexpected results when strong teams lose away to weak teams. Here we examine the degree to which the home field advantage is really an advantage for various types of sports and propose a mathematical framework to explain the observed differences.

First we considered soccer which undeniably is the most popular sport on the planet. We chose the English Premier League (EPL) and for a given season and from all teams we found the total number of wins at home and total number of wins away (the data were downloaded from www.soccerbase.com). For a given season we define the ratio of the total number of wins at home to the total number of wins away as *r*. The second column in table 1 shows *r* for eight consecutive seasons. The average of these values (n = 8) is $\bar{r} = 1/n\sum_{i=1}^{n} r_i = 1.75$ and the standard deviation is $s = \left[\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2\right]^{1/2} = 0.15$. This average value indicates a strong home field advantage as on the average a team wins 1.75 times as many games at home than away.

The question we asked next was whether such an advantage applies to other sports. In order to address this question we considered the National Football League (NFL), the National Basketball Association (NBA), and the Major League Baseball (MLB) in USA and repeated the above analysis. Note that in this analysis we avoided using other leagues (like the Major League Soccer and the National Hockey League in USA) where tied games are broken by either penalty kicks or sudden death. This introduces a bias and may affect the conclusions. The data were obtained from the archives of the Milwaukee Journal and the sites www.nfl.com, www.nba.com, and www.baseball.com. Columns 3, 4, and 5 in table 1 show the results for NFL, NBA, and MLB respectively. Note that (1) in 1994 the MLB players went on strike and the season was aborted, and (2) the regular season for NFL is in the first year of the two year period indicated in column 1 and for MLB the season is in the second year. Table 1 shows the results. We find that for NFL: $\bar{r} = 1.50$, $s = 0.15$, for NBA: $\bar{r} = 1.51$, $s = 0.10$, and for MLB: $\bar{r} = 1.16$, $s = 0.033$. Do these results indicate statistically significant differences in home field advantage in the different sports?

In order to answer this question we proceeded as follows. We assumed that *r* is normally distributed with a mean μ and variance σ^2 and that the estimated standard deviations represent the actual variances (i.e. $s^2 = \sigma^2$). Then, for any two different types of sports we constructed the null hypothesis $H_0: \mu_1 - \mu_2 = \delta$ against the alternative $H_1: \mu_1 - \mu_2 \neq \delta$. If the absolute value of the statistic $z = (\bar{r}_1 - \bar{r}_2 - \delta) / \left[\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right]^{1/2}$ is less that $Z_{\alpha/2}$ where n is the sample size (equal to eight except for MLB where it is equal to seven), α is the level of significance, and z is the corresponding value of the standard normal distribution, then the null hypothesis cannot be rejected. Otherwise it is rejected. Considering $\delta = 0$ (i.e. there are no differences between the two means) we find that the null hypothesis is rejected at a significance level of 0.05 for any two of the four sports considered here, except between American basketball and football. This means that the observed differences in home field advantage are statistically significant except between NBA and NFL. Note that if we relax the assumption that the estimated variances represent actual variances, and instead consider testing the null hypothesis H_0 against the alternative H_1 assuming either unknown unequal variances or unknown equal variances (in both cases the t distribution is used but the degrees of freedom are computed differently), we again find that the observed differences in home field advantage are statistically significant at a significance level of 0.05 expect between NBA and NFL. Note also that by comparing the results in table 1 to the case of no home field advantage at all (i.e. $\bar{r} = 1$, $s = 0$) we find that in all four sports the reported $\bar{r}$ values are significantly different from 1. This means that home field advantage is a characteristic of all types of sports but it varies from one type of a game to another, with soccer exhibiting the strongest home field advantage and baseball exhibiting the weakest.

Since in any given sport both teams are subject to exactly the same rules we suggest that home field advantage is a manifestation of the effect of increased psychological stress for the visiting team compared to the home team. Then, we hypothesize that in sports with lower home field advantage this stress *difference* between the home and the visiting team is somehow reduced. Accepting this conjecture and considering the nature of these games we propose the following mathematical interpretation of home field advantage.

If we consider a game as an N-element (plays) sequence then the information entropy is defined as $H(s) = -\sum_{i=1}^{N} P_s(s_i) \log P_s(s_i)$

Table 1

Season	r, EPL	r, NFL	r, NBA	r, MLB
1992–93	1.81	1.58	1.53	1.19
1993–94	1.47	1.23	1.58	–
1994–95	1.67	1.31	1.45	1.14
1995–96	1.95	1.51	1.54	1.16
1996–97	1.64	1.62	1.38	1.17
1997–98	1.82	1.54	1.36	1.16
1998–99	1.76	1.69	1.65	1.09
1999–00	1.85	1.48	1.57	1.18
$\bar{r}$	**1.75**	**1.50**	**1.51**	**1.16**
s	**0.15**	**0.15**	**0.10**	**0.033**

where $P_s(s_i)$ is the transition probability that a measurement of the state s yields s_i. To understand the relevance of this equation is helpful to think of entropy as a measure of the uncertainty associated with the measurement of a state or the "quantity" of surprise we feel when we read the result of a measurement. Unexpected (low-probability) measurements carry greater entropy than do expected (high probability) measurements. In order to minimize unexpected events we must, therefore, keep the game at a low entropy level. This corresponds to keeping the game at a high amount of information level[1]. Since in any evolution the initial amount of information (which in a game can be assumed to be the initial game plan) is being gradually lost, in order to keep a high information level information must be constantly transferred effectively. Transferring information during a game depends on the number of designed plays and the effectiveness in delivering and executing each design. Delivering designed plays depends on the degree of communication between players and coaches, which in turn depends on the continuity of the game, the number of substitutions and time-outs, and the signaling of the plays. Executing the designed plays depends among others on the imposed time constraints between plays. When we look at the rules and dynamics of each of the considered sports we see that in baseball not only all plays are designed and discussed with each player but there are no time constraints in completing a designed play. As such information is transferred effectively throughout the game which according to the discussion above should minimize unexpected plays. In soccer, which is a highly continuous game, the only designs (apart from the initial pre-game plan) are the half-time discussions and adjustments from only three substitutions. Accordingly, information transfer is not very effective and the chances for unexpected results increase. In football and basketball most of the plays are designed, there are unlimited substitutions and many time-outs but in both there are time constraints in completing the plays. Consequently, as information transfer is concerned, they are very similar and as such they fall in between baseball and soccer.

A measure that relates to information transfer in a game could be defined as the ratio, ρ, between the preassigned length of the game (48 minutes for basketball, 60 minutes for football, 90 minutes for soccer) and the average length of time that a game actually lasts. If there is no stopping the time, the game is very continuous, information transfer is little, and the ratio is close to one. If there are a lot of breaks between plays, information transfer increases and the ratio decreases. For baseball there are no time constraints at all and as such we assume that the information transfer is maximized (i.e. $\rho \to 0$). For the other sports we find that ρ 0.30, 0.33, and 0.95 for basketball, football, and soccer, respectively. These values are consistent with the values of the ratio r and provide support for our proposed interpretation of home field advantage.

Within this framework we, therefore, find consistency between the above reported home field advantage values and amount of information transfer. A simple explanation for this relationship may be that signaling plays and frequent breaks and substitutions, which result in designed plays (i.e. increased information transfer), reduce the stress difference between home and visiting team, thereby increasing the likelihood of an away win. Understanding home field advantage in the above terms provides the means to explore new strategies for information transfer in a game, which will increase the chances for an away win.

Acknowledgement

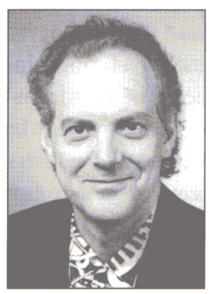
We would like to thank an anonymous reviewer for his suggestions and comments.

REFERENCE

1. Wiener, N. *Cybernetics* (MIT Press, Cambridge, Massachusetts, 1961).

ANASTASIOS A.TSONIS
UNIVERSITY OF WISCONSIN-MILWAUKEE, USA
DEPARTMENT OF MATHEMATICAL SCIENCES

PANAGIOTIS A.TSONIS
UNIVERSITY OF DAYTON, USA
DEPARTMENT OF BIOLOGY



Anastasios A. Tsonis is a Full Professor in the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee. He holds a B.Sc degree in Physics and Mathematics from Aristotle University of Thessaloniki, Greece and M.Sc and Ph.D degrees in Atmospheric Sciences from McGill University, Montreal, Canada. His main research interest is nonlinear dynamical systems and their applications to climate variability and dynamics. He is the author of more than 70 peer reviewed papers and three books. **Panagiotis A. Tsonis** is a Full Professor of Molecular Biology in the Department of Biology, at the University of Dayton. He received his B.Sc from the University of Patras, Greece, and his M.Sc and Ph.D degrees from Nagoya University, Japan. His main research interest is biology of tissue regeneration. He is the author of nearly 100 papers and two books. Apart from their main interests the two brothers work together in topics such as bioinformatics and the mathematical structure of DNA sequences where they have co-authored 10 papers.