JoE

Dominance, Streakiness, and Non-random Patterns in the Game of N!àì

Yi Lu[1], Alex de Voogt[2], and Fernand Gobet[3]

[1] Department of Mathematics and Computer Science, Drew University, USA
[2] Department of Business, Drew University, USA
[3] Centre for Philosophy of Natural and Social Science, London School of Economics (LSE), UK

Correspondence: Fernand Gobet, F.Gobet@lse.ac.uk

Journal of Expertise
2024. Vol. 6(4) & 7(1)

ISSN 2573-2773

Abstract

In the study of play, it has been suggested that hunter-gatherer and egalitarian societies avoid competitive games and forms of dominance in society. The game of N!àì is played by the Ju|'hoan people in Botswana in a way that facilitates competition and contradicts this suggestion. While similar games have been described and studied extensively from qualitative perspectives, we aim to analyze patterns in the game of N!àì from a quantitative and statistical perspective. Specifically, using original data extracted from first-hand video recordings of twenty different matches, we adopt a class of Bradley-Terry pairwise comparison models to analyze players' latent game-playing abilities. In addition, we demonstrate the use of a Bayesian segmentation model to quantify streakiness. The assessment of the level of streakiness further enables the discussion of whether it is more advantageous to play in a streaky or in a random fashion. The findings question more general assumptions about hunter-gatherer societies and the activities in which they are likely to engage due to the egalitarian nature of their society. Implications for expertise research beyond gesture games are discussed.

Keywords

Bradley-Terry, paired comparison, Bayesian, streak, game, dominance

Introduction

The literature on expertise in games has been dominated by board games, in particular chess, mostly due to the presence of rating systems and regular tournaments (Gobet, de Voogt, & Retschitzki, 2004). Board games played in Asia and Africa have supplemented the research results; these games mostly consist of mancala games, such as bao (de Voogt 1995, 2002) and awèlé (Retschitzki 1990, N'Guessan, 1992). In this study, we introduce an African game, N!àì, which has no tournament or rating tradition. It is, however, part of an on-going discussion of whether hunter-gatherers play or develop strategic games. Specifically, we quantitatively evaluate dominance in the game from a statistical perspective.

Games that are played in naturalistic settings, without tournaments and rating systems, are rarely used for research in expertise as they do not allow a convenient measure to distinguish experts from novices. Any attempt at doing so would require large numbers of matches between players to generate or approximate a ranking of players. For example, a tournament for the game of bao was organized in 1994 to allow for descriptive statistics of the game and it featured 92 matches for ten players in the period of one month (de Voogt, 1995). In chess, one player needs to play at least 25 games before they can receive an official Elo-rating (https://new.uschess.org/frequently-asked-questions-member-services-area), suggesting that even a month of documenting matches is not nearly enough for establishing a rating system (see also Aldous, 2017).

The games of Morra and Rock-Paper-Scissors stand out as games for which skill and expertise are documented even though they are not board games and do not have a rating system or recognized championships. Instead, they are mostly perceived as games that generate random outcomes. However, expert players have been identified in those games, not the least because of the constraint imposed by the speed of play: A match between two players takes only seconds. Documenting large numbers of interactions between opponents allows for the necessary data to analyze the actions of the players and determine the existence of dominant or statistically superior players.

Like Morra and Rock-Paper-Scissors, the game of N!àì falls in the category of gesture games. In these games, two players gesture with a hand or a set of fingers at a precise moment in time, where the combination of each player's signal determines the winner. These games have offered "a unique research paradigm to study sequential adversarial strategies in repeated interactions" (Zhang et al., 2021). It is hypothesized that the optimal strategy in these games is to play purely randomly; "however, humans are subject to the influence of many factors such as physical environment, emotional state, and degree of experience which ensure the game's outcome is variable" (Serra, 2020, p. 24). It is suggested "that Morra is hardly a game of luck. An expert player is able, in fact, to produce effective strategies which lead almost invariably to success against less skilled players" (Delogu et al., 2020, p. 2). In the case of Rock-Paper-Scissors, human strategies have been categorized from different perspectives (Dyson, 2019), and winning strategies have been researched in detail (Alfaro et al., 2009).

Despite its similarities to the other gesture games, the game of N!àì has so far not been used for research on expertise, strategy, or randomized play. There are no championships, and the game has not been simulated with statistical models (de Voogt, 2017). The possible presence of experts would contradict the notion that, by definition, N!àì is a game of luck since the players of N!àì are said to live in an egalitarian society in which, according to some scholars (Roberts et al., 1959; Chick, 1998), a strategic game with dominant players would be disruptive. The game is played frequently, often in teams as large as five, and suggestions have been made in the literature that some players are better than others, even though the players themselves maintain that no player dominates the game (de Voogt, 2017). (It is noted that the terms "dominant" and "dominance" are used here in their anthropological sense rather than the specific meaning they have in a game theoretical context; the phrases "dominant player" and "superior player" are used here interchangeably.)

The presence of large teams, compared to Morra and Rock-Paper-Scissors, possibly complicates a player's perception regarding individual dominance in the game but no statistical analysis of N!àì outcomes at the individual or team level has so far been attempted. To study the characteristics of the game as played in its current cultural context as opposed to an experimental or simulated environment, we collected videos of people playing this game in Botswana for a month-long period. The results of this study should inform to what extent N!àì can be considered a game of luck or expertise, and whether its non-random aspects could be salient to the participants in the game.

Statistically, the basic structure of the game of N!àì can be regarded as pairwise comparisons between two individual players. Bradley-Terry style models (Bradley & Terry, 1952) have been widely used in similar setups to assess player or team superiority in sports (for example, chess, Elo, 1978; golf, Baker & McHale, 2016; soccer, Schauberger et al., 2018; and tennis, McHale & Morton, 2011). In this work, we use this model to compare the latent game-playing abilities of each player and evaluate the hypothesis that N!àì, like Morra and Rock-Paper-Scissors, is not entirely a game of luck and that some players have demonstrated superiority in the game, especially against certain opponents. We also aim to identify non-random patterns by looking at winning and losing streaks. We adopt a Bayesian segmentation model that has been proposed to quantify streakiness (Yang, 2004). We demonstrate a few limitations of this method when applied to our data and augment it by adding another metric of significance based on empirical *p*-values calculated via Monte Carlo simulations. While streakiness has typically been discussed in sports in the context of winning or losing streaks, in our application, levels of streakiness can also be used to evaluate a player's preferred strategy by detecting streaks of left- or right- hand plays. By investigating to what extent such a strategy affects the outcomes of the game, we assess the hypothesis that the optimal strategy is to play the game as randomly as possible.

Rules of N!àì

Rules of the game of N!àì have been detailed in a previous study (de Voogt, 2017). In summary, during each game play, two opposing players simultaneously play one hand on a particular beat in the accompanying music, and the winner is determined by whether the two players played the same or different hands. One player is on the team whose winning condition is "same hands" and wins if the two players played the same hands (i.e., both left or both right); the opponent player is then on the team whose winning condition is "different hands" and wins if the two players played different hands (i.e., one left and one right, or vice versa). The winning conditions are arbitrarily assigned beforehand by the two teams that participate in the game, and no player (or team) is specifically linked to one of the winning conditions.

Participants, Recordings and Summary Statistics

Video recordings of the game were collected over a one-month period in a setting with ten players, seven men and three women, who regularly demonstrated the game for guests of a safari lodge in Botswana. Data extracted from the videos include the specific individuals participating in the game, the position of players in their team line-up, the order of each play, the outcome of each play as well as the hand that each player used for the play. The recordings were made by one of the authors with permission of the players and the venue where the players were working, and the process followed procedures set out by the Institutional Review Board of the American Museum of Natural History in New York. In total, data for 20 different game days were obtained, with each game day corresponding to one match between two different teams. A summary of the 20 matches is shown in Table 1. For instance, during the first match, a total of 7 rounds were played with team 1 winning 3 rounds and team 2 winning 4 rounds. To clarify our terminology, at the player level, a player wins a *point* if he or she wins against the opposing player. When the players on a team have reached a total of five points, this team wins a round and the team's *score* increases by one.

The players preferred certain team members over others due to a developing rivalry between two groups of five. As shown in Table 1, team 1 refers to the team that consists mostly or exclusively of players A, B, C, D, or E, and team 2 refers to the team that mostly features players F, G, H, I, or J. Only when players were not available would a player switch sides to even the size of the two teams. If for a specific match a team has all of the regular five players or is missing only one of them, we consider it an "ideal team" (see Table 1). This variable is used in later sections to assess whether or not a player's game-playing ability is affected by that person's team members.

The duration of each play varied from about 4 seconds to over 8 seconds per interaction. For instance, two players, J and E, particularly enjoyed prolonging their encounters by foregoing the beat in the music repeatedly, feigning and using entire body movements rather than just their hands.

Table 1. Team level statistics for the 20 recorded match days. The score of the winning team is highlighted in bold.

	Team 1			Team 2		
Match	Score	Players	Ideal team	Score	Players	Ideal team
1	3	A B C D E	Y	**4**	G H I J	Y
2	4	A B C D E	Y	4	G H I J	Y
3	**12**	A B C D E	Y	9	G H I J	Y
4	3	A B C D E	Y	**8**	G H I J	Y
5	**5**	B C D E	Y	2	F G H I J	Y
6	**20**	A B C D E	Y	13	F G H I J	Y
7	**8**	A B C D E	Y	7	F G H I J	Y
8	0	A B C D E	Y	**6**	F G H I J	Y
9	3	A B C D E	Y	**5**	F G H I J	Y
10	**6**	B C E	N	4	F G H I	Y
11	6	A B C E	Y	**7**	F G H I J	Y
12	**9**	A B C E	Y	8	F G H I J	Y
13	**13**	A B C E	Y	11	F G H J	Y
14	**3**	B H	N	1	D E	N
15	**2**	B I	N	1	D E	N
16	**4**	A C	N	0	J	N
17	**7**	A C D E	Y	4	F G H J	Y
18	7	A B C D E	Y	**10**	F G H I J	Y
19	**5**	A B D	N	2	F G H J	Y
20	**7**	B C E F	N	4	G H I J	Y

Method

Randomness itself cannot be directly observed; only departures from randomness, such as a particular order, can be quantified (Towse & Neil, 1998). We aim to explore the level of randomness in both the outcome and the strategy of each play in order to determine the salience of superior play in the game as well as to identify playing strategies. Specifically, in our approach, we first adopt a Bayesian segmentation model to analyze and quantify streakiness (a higher presence of wins or righthand plays in a row). We carry out a simulation study to demonstrate limitations of this approach and augment it with a Monte Carlo procedure and empirical p-values. We then evaluate player dominance with the classic Bradley-Terry model for paired comparisons. Variations of this model also allow the evaluation of factors that may affect a player's latent game-playing ability.

Models for Streakiness

To identify streakiness in each player's performance, we adopt a Bayesian segmentation model (Yang, 2004). In general, given data $D \equiv y_1, \dots, y_n$, which are assumed to be independent binomial random variables with parameters $(m_i, p_i), i = 1, \dots, n$, we can estimate whether or not there exists a cut point $c \in \{1, \dots, (n-1)\}$, such that the event probabilities $p_i = p_1$ for $i \leq c$ and $p_i = p_2$ for $i > c$. The null model $M_0\colon p_1 = p_2$ (i.e., no streak) is tested against the one-cut-point model $M_1\colon p_1 \neq p_2$ and the model selection criterion is based on the Bayes factor of model M_1 against M_0, denoted by B_{10}. Detailed derivations of B_{10} can be found in the Appendix.

Simulation Study and Empirical *p*-Values

To decide between models M_0 and M_1, it has been suggested simply to consider if $B_{10} > 1$ (Yang, 2004). We design a simulation study to explore properties of this criterion, in terms of its false positive rate and power. To evaluate the false positive rate, we simulate a binary data sequence from a true underlying model M_0 (no cut point) with event probability p_0. We then calculate the value of B_{10} for the simulated data sequence; if $B_{10} > 1$, we would (falsely) select model M_1. We repeat this process for a large number of times and calculate the proportion of data sequences that yield a B_{10} value greater than 1, which gives an empirical estimate of the false positive rate. We test a few different combinations of the sample sizes ($n \in \{10, 20, \ldots, 100\}$) and event probabilities ($p_0 \in \{0.5, 0.55, 0.8\}$). Results are shown in panel (a) of Figure 1.

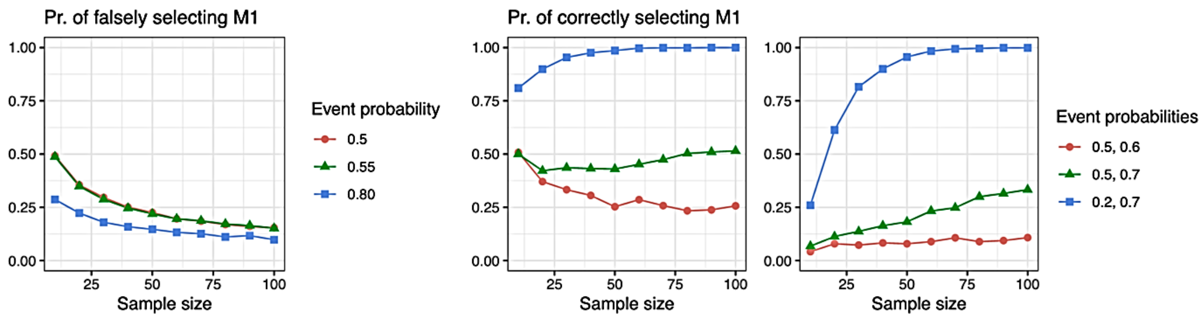

Figure 1. (**a**) False positive rate (left panel), (**b**) power based on B_{10} (middle panel), and (**c**) power based on p-value by varying sample sizes and event probabilities. For the false positive rate, the event probability refers to the value of p_0 under the no-streak model. For the power, the event probabilities refer to the values of (p_1,p_2) under the one-streak model. Sample size values are 10,20,...,100.

We see that the false positive rate tends to decrease when the sample size increases, as expected. The false positive rate is also comparatively smaller when the event probability differs from 0.5 by a large margin, since, in that case, data favor model M_0 with more certainty toward the constant event probability (i.e., the sample proportion has smaller variance). For our application, however, the event probabilities are typically estimated to be between 0.4 and 0.6, and the sample size ranges from 4 to 63, which would result in false positive rates roughly between 20% and 50%.

To remedy the issue of high false positive rates when the selection criterion is only to consider if $B_{10} > 1$, we instead treat B_{10} as a test statistic and calculate empirical p-values based on Monte Carlo simulations. Specifically, suppose the observed data $D_{obs} = y_1, \ldots, y_n$ yields a B_{10} value $B_{10}^{(obs)}$. For each simulation step k, we simulate n binary outcomes $D^{(k)} = y_1^{(k)}, \ldots, y_n^{(k)}$ with no cut point and with event probability being the sample proportion of 1's in the observed data D_{obs}; the corresponding B_{10} value is denoted by $B_{10}^{(k)}$. We repeat this process for a large number K times to obtain $\{B_{10}^{(1)}, \ldots, B_{10}^{(K)}\}$. Intuitively, these values provide an estimate of the distribution of B_{10} under the null model M_0. Let $S = \sum_k I(B_{10}^{(k)} \geq B_{10}^{(obs)})$ be the number of simulated values that are greater than or equal to the observed value; the empirical p-value is then calculated as $\frac{S+1}{K+1}$ (North et al., 2002). Using the empirical p-values, we can control the false positive rate to be 5%, for example, by performing the test at $\alpha = 0.05$ level.

Power of the aforementioned model selection criteria (based on the B_{10} value alone or based on the p-value at $\alpha = 0.05$ level) can be evaluated by simulating binary data sequences from a true underlying model M_1 (one-cut point) with event probabilities (p_1, p_2) and cut point at 30% of the sample size. Results are shown in panel (b) of Figure 1. We see that both criteria tend to have low power (< 50%) when the event probabilities before and after the cut point are moderately close ($p_1 = 0.5, p_2 = 0.6$; or $p_1 = 0.5, p_2 = 0.7$), regardless of the sample size and the selection criteria. The result seems reliable only when the event probability changes drastically ($p_1 = 0.2, p_2 = 0.7$) and the sample size is at least 50. Intuitively, a large amount of binary data is needed to detect when the success probability changes by a small amount. The issue of low power is difficult to avoid with limited data and we acknowledge that the proposed procedures are unlikely to detect small changes in the event probability, even if a cut point does exist.

Models for Pairwise Comparison

To assess if one player is superior to another player, we adopt the classic Bradley-Terry model for paired comparisons (a comprehensive review can be found in Cattelan, 2012). The basic form (referred to as an unstructured model) is given by

$$Y_{kij} \sim (indep)\ Bern(p_{kij}),$$
$$logit(p_{kij}) = \mu_i - \mu_j,$$

where $Y_{kij} \in \{0,1\}$ is the outcome of the *k*th game between player *I* and player *j* (1 if player *i* wins against player *j*, 0 otherwise), and μ_i and μ_j are *worth parameters* that represent the game playing abilities of player *i* and *j*. This model can be modified to include *p* explanatory variables $x_1, \dots, x_p$, as well as a zero-mean random effect ξ. In this case (referred to as a structured model), the worth parameter for player *i* is expressed in the linear form

$$\mu_i = x_{i1}\beta_1 + \cdots + x_{ip}\beta_p + \xi_i.$$

Models with only fixed effects assume that the game plays are independent from each other, whereas the random effect model assumes that game plays involving the same player are correlated. In addition, we can add a game-specific "advantage" term z_{kij}, where $z_{kij} = 1$ if player *i* has the advantage in game *k* and $z_{kij} = -1$ if player *j* has the advantage (for instance, in sport applications, this term can represent a home court advantage). In this case, we have

$$logit(p_{kij}) = \mu_i - \mu_j + \beta \cdot z_{kij},$$

where the parameter β represents the size of the advantage and the worth parameters represent the game-playing abilities in the absence of this advantage (Turner & Firth, 2012).

Results

Player-level summary statistics are provided in Table 2. Standard one-sample test for proportions (with *p*-values displayed in Table 2) show that Players B and C are significantly more likely to win than to lose (winning percentages are 57.2% and 55.5%, respectively), and Player I and J significantly prefer to play one hand over the other (right hand percentages are 57.7% and 43.8%, respectively). In addition, it is shown that player B played left and right hands in nearly equal measure, suggesting that this strategy may be the most effective. This "evenhandedness" appeared in individual matches as well as in all matches combined. In contrast, at the team level, no dominant performance could be determined. In total, disregarding matches 14, 15, and 16 due to a lack of preferred team members, team 1 won 10 out of 17 matches (58.82%) with a total score of 118, compared to a total score of 108 for team 2. However, a preliminary one-sample test for proportions shows that team 1's better performance is not significant.

We apply the Bayesian segmentation model to the player-match level outcomes and calculate the corresponding empirical *p*-values to further detect the existence of nonrandom patterns in terms of streakiness. We then apply different variations of the Bradley-Terry model to see if some players have significantly better game-playing abilities. We consider several covariates in the model, to answer questions such as whether a player's playing ability is affected by his/her team members, and whether it is better to play randomly (i.e., switching between a left and a right hand with no pattern) or with some strategy in mind (i.e., playing a right hand in a streak).

Evaluating Streakiness

Since the 20 matches are played on different days and under different conditions (such as length of the match and team size), we only consider streakiness within each match. We assess two types of streakiness by looking at whether a player has a sequence of wins or losses, which is an indication of the salience of a player's dominance, or a sequence of left- or right-hand plays, which suggests possible nonrandom playing strategies. Data D_{im} consist of binary outcomes for player i ($i \in \{A, B, \ldots, H\}$) in match m ($m \in \{1, \ldots, 20\}$). We calculate the value of B_{10} for each set of data D_{im}, as well as an empirical p-value based 10,000 Monte Carlo simulations. Results show that 5 sets of data sequence display a significant streaky pattern in terms of game results, shown in panel (a) of Figure 2. For each of the streaky data sequences, we also show the estimated cut point (estimated using the posterior mean). For instance, we can see that player A played a total of 9 games in match #16, starting with a streak of 5 wins, and then lost 3 out of the 4 remaining games. On the other hand, we also detect 9 sets of data sequence that are significantly streaky in terms of which hand is played, shown in panel (b) of Figure 2. For instance, player A played a total of 52 games in match #6. Seemingly, there is a change in strategy around game 20 (the cut point is estimated to be 18.8): this player played left and right hands fairly evenly before the cut point but then had streaks of playing a left hand after the cut point.

Table 2. Winning percentage for each player and how often they played their right hand, with p-value for testing if the percentage is significantly different from 0.5 in parenthesis. Significant values are highlighted in bold.

Player	Number of plays	% win (p-value)	% right hand (p-value)
B	442	**0.572 (0.003)**	0.495 (0.887)
C	425	**0.555 (0.026)**	0.480 (0.438)
I	326	0.503 (0.956)	**0.577 (0.007)**
D	256	0.500 (1.000)	0.453 (0.151)
J	363	0.499 (1.000)	**0.438 (0.021)**
F	268	0.496 (0.951)	0.506 (0.951)
G	422	0.472 (0.263)	0.543 (0.088)
H	408	0.471 (0.255)	0.532 (0.216)
A	300	0.457 (0.149)	0.445 (0.074)
E	382	0.453 (0.073)	0.524 (0.384)

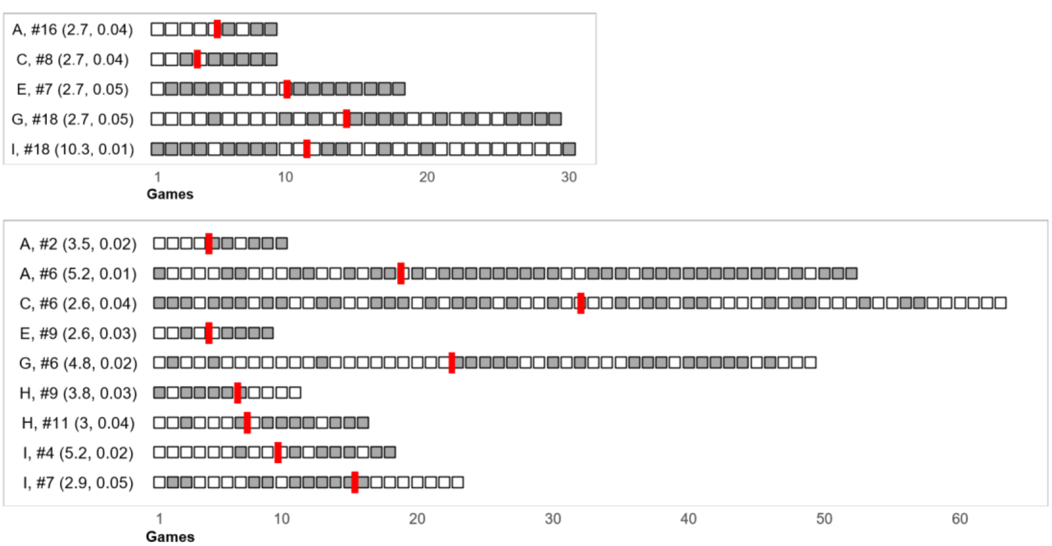

Figure 2. Streaky data sequences in terms of (**a**) game result (top panel) and (**b**) left/right hand (bottom panel). For each data sequence, we show the player, the match number, and (in parenthesis) the B_{10} value and the empirical p-value. For the top panel, white (grey) squares correspond to wins (losses); for the bottom panel, white (grey) squares correspond to right-hand (left-hand) plays. The red line in each data sequence corresponds to the estimated cut point.

While we have detected patterns of streakiness that are considered unlikely due to random chance at the individual player-match level, we caution against drawing a conclusion that the game of N!àì, in general, is non-random in terms of streakiness, due to the issue of multiple comparison. In fact, the percentage of significant data sequences D_{im} (3.1% for game results and 5.6% for left/right hand) are close to what one would expect to happen by random chance when the tests are performed at $\alpha = 0.05$ level. On the other hand, we note limitations of the procedure: It is unlikely that a player's winning percentage or right-hand percentage changes dramatically even in the presence of a cut point, and the test based on B_{10} and its empirical *p*-value likely does not have enough power to detect an existing streak in this case, as discussed previously. This leads to high false negative rates, and the overall level of streakiness in the game is likely underestimated.

Evaluating Player Superiority and Optimal Strategy

To evaluate the hypothesis that some players are significantly better than others and are especially superior against certain opponents, we first fit the unstructured Bradley-Terry model. To avoid issues of identifiability, we choose player E (whose winning percentage is the lowest) as the baseline category. Estimates of the worth parameters for each player and their standard errors are given in Table 3.

As expected, all of the estimates are positive, and the estimated game-playing abilities of players B and C are especially superior. Furthermore, we estimate the pairwise Tablecontrasts $\mu_i - \mu_j$ between all pairs of players and generate 95% confidence intervals based on the (approximate) normal distribution. After an exponential transformation, we obtain confidence intervals for $\frac{p_{ij}}{1-p_{ij}}$, the odds that player *i* wins against player *j*, which is shown in Figure 3. Based on these results, we see that players B and C are significantly better than players E, A, H, and G. When player B plays against player E, for instance, the odds of winning is estimated to be around 1.6, corresponding to an estimated probability of winning $\widehat{p_{BE}} = 61.5\%$.

We fit additional models that seek to further explore what factors contribute to a player's game playing ability. These models are fitted using the R package *bradleyterry2* (Turner & Firth, 2012). We first add a team advantage term, where a player is considered to have an advantage if they are in an "ideal team" (as summarized in Table 1). Results show that this perceived team advantage is not significant (coefficient estimate is -0.055 with a standard error of 0.138 and *p*-value of 0.689). The estimated worth parameters, given in Table 3, are similar to the model without the team advantage term.

We also quantify how evenly and randomly a player tends to play left and right hands and use this information as covariates for the player's worth parameter. Specifically, for each player, we look at the corresponding left/right hand B_{10} values and calculate the average of these values as a "streak score." Intuitively, players with a higher streak score tend to play in a streakier fashion. We also calculate a "hand preference score" as $|\widehat{p_{rh}} - 0.5|$, where $\widehat{p_{rh}}$ is a player's overall right-hand percentage (given in Table 2). This score gives a simple measure of how much a player favors a specific hand. We fit two separate fixed effect Bradley-Terry models using the streak score and the hand preference score as covariate. Note that we avoid putting these two variables in the same model because they are highly correlated (correlation = 0.715). This is expected because, for instance, a very high right-hand percentage is likely linked to streaks of righthand plays. Coefficient estimates are −0.352 (S.E. = 0.165, *p*-value = 0.033) for the streak score and −2.52 (S.E. = 1.37, *p*-value = 0.065) for the hand preference score, respectively. This implies that the estimated worth parameters will be higher for players who have low streak scores and low hand preference scores. Overall, these results support the hypothesis that the optimal strategy is to play as randomly as possible.

Table 3. Estimates of the worth parameters for each player, as well as the standard errors of the estimates, based on (a) the unstructured model and (b) model with a team advantage term. Player E is used as the baseline when fitting both models.

	(a)		(b)	
Player	Estimate	S.E.	Estimate	S.E.
B	0.465	0.139	0.463	0.139
C	0.403	0.142	0.404	0.142
I	0.261	0.144	0.266	0.145
J	0.222	0.140	0.228	0.141
F	0.207	0.154	0.209	0.154
D	0.194	0.163	0.200	0.163
G	0.127	0.132	0.139	0.135
H	0.116	0.135	0.125	0.138
A	0.007	0.155	0.011	0.156
E	0	0	0	0

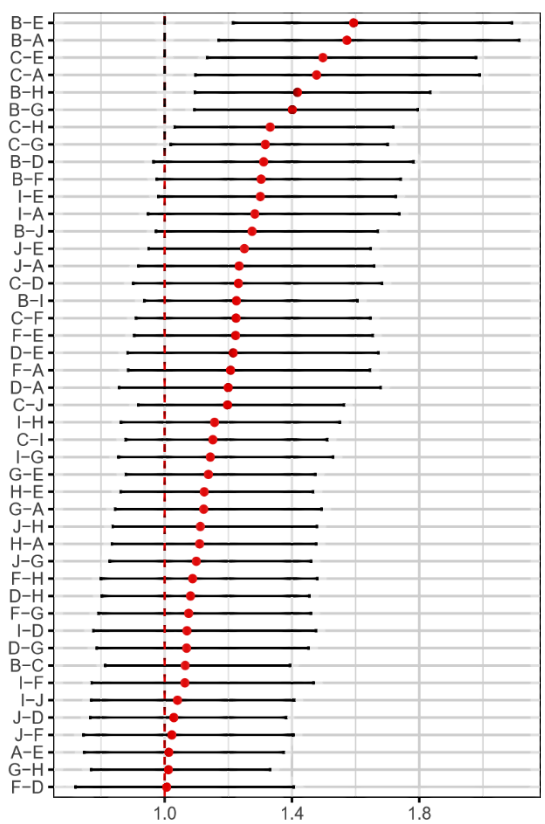

Figure 3. 95% confidence intervals for the odds of winning between pairs of players. For instance, the top line corresponds to the interval estimate of the odd of player B winning against player E, given by $\exp(\mu_B - \mu_E)$.

Discussion

The results of the analysis of N!àì only partly support the expectations based on previous analyses of Morra and Rock-Paper-Scissors. While our hypothesis that some players may dominate the game is confirmed, only two players showed consistent superiority while the remaining eight players displayed no significant differences between their performances in the game. In other words, there is no ranking of players based on significant differences apart from the two top players. Moreover, at the team level—an aspect not present in the same way in Morra and Rock-Paper-Scissors—dominance was not significant.

Based on the pairwise comparison models, individual superiority is found between the top two players and four other players only (see Figure 3). Two of these lesser players, E and A, who are the worst two, appear to be of the same team as the two top players B and C. It means that the top two players got to play the bottom two players much less often and that at the team level dominance would only be visible against two opposing players and not against the other three. As a result, the composition of the teams has made the presence of superior players much less visible. This makes the teams especially competitive, but it does not suggest to individual players that any one of them dominates the game more than others.

The presence of winning streaks, an element of the game that is arguably salient to individual players, is detected at the individual level. The overall frequency of streaks, however, does not appear to be significantly higher than expected in a random match. Those players who showed a high streakiness of the left or right hand were not more successful, suggesting that these streaks, if salient, were not likely a part of a winning strategy. We note that more data are needed to improve the power of the proposed procedure.

The absence of dominance at the team level and the limitation on winning streaks obscure the salience of individual dominance in the game. Conversations with the players confirm that superiority is not noticeable although at the same time, they agree that in a particular match some player or one team may significantly outscore another. It has been repeatedly stated in the literature that the best strategy in these types of games is random play and that environmental factors preclude players from achieving such an optimal strategy. The superior players in our data set, compared to their fellow players, came particularly close to "evenhandedness" or the same frequency of using the left or right hand both during a match and across all games combined. The conviction that the game is one of luck may have inspired these two players to select their hand of play without any particular strategy or close to randomly, which ironically has made them dominate the game over time.

Future research may explore other possible strategies employed by players in this data set as well as the role of competitive play in Ju|'hoan and other San societies, which do not seem to recognize superior players but still foster intense competition. The analytical methods we have employed for the game of N!àì, in particular the presence of streakiness, may also be applied to the games of Morra and Rock-Paper-Scissors. Streaks of using the same gesture or winning streaks in these games are open to the same method of analysis explained in this study. This would further our understanding of their playing strategies and broaden the comparison between these types of games.

Beyond gesture games, the statistical methods we have used will be of interest in the analysis of data of experts in naturalistic settings. By definition, experts are rare and the data they produce suffer from a number of shortcomings – just like in the game of N!àì discussed in this article. Experts' data can sometimes be approximated by collecting data from non-experts, typically undergraduate students. However, it is often preferable to use imperfect ecological, naturalistic data produced by few experts rather than statistically more suitable data produced by many non-experts. In these cases, the techniques used in this article can shed important light on expert behavior.

Authors' Declarations

The authors declare that there are no personal or financial conflicts of interest regarding the research in this article.

The authors declare that the research reported in this article was conducted in accordance with the Ethical Principles of the *Journal of Expertise.*

The authors declare that the dataset is not publicly available but can be provided upon request.

Acknowledgements

This study would not have been possible without the help of Rebecca Rivera and Mohamed Ibrahem in the video analysis of the game of N!àì. We are grateful for the support of Richard Butler, Elizabeth DeGaetano, Michael Turner, and the American Museum of Natural History. Finally, we thank the staff and management at Jack's Camp, Botswana, especially Eugene Khumalo and Ralph Bousfield, for their generosity and kindness, but, most of all, the Ju|'hoan people, whose enthusiasm for this game has made this research such a pleasure.

ORCID iDs

Yi Lu
https://orcid.org/0000-0003-4528-2665

Alex de Voogt
https://orcid.org/0000-0002-1763-2288

Fernand Gobet
https://orcid.org/0000-0002-9317-6886

References

Aldous, D. (2017). Elo ratings and the sports model: A neglected topic in applied probability? *Statistical Science, 32*, 616-629.

Alfaro, R., L. Han, and K. Schilling (2009). Winning at rock-paper-scissors. *The College Mathematics Journal 40*(2), 125–128.

Baker, R. D. and I. G. McHale (2016). An empirical Bayes' procedure for ranking players in Ryder cup golf. *Journal of Applied Statistics 43*(3), 387–395.

Bradley, R. A. and M. E. Terry (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika 39*(3/4), 324–345.

Cattelan, M. (2012). Models for paired comparison data: A review with emphasis on dependent data. *Statistical Science 27*(3), 412–433.

Chick, G. (1998). Games in culture revisited. *Cross-Cultural Research 32*(2), 185– 206.

de Voogt, A. (1995). *Limits of the mind: towards a characterization of bao mastership*. Leiden: CNWS Publications.

de Voogt, A. J. (2002). Reproducing board game positions: Western chess and African bao. *Swiss Journal of Psychology, 61*, 221-233.

Delogu, F., M. Barnewold, C. Meloni, E. Toffalini, A. Zizi, and R. Fanari (2020). The Morra game as a naturalistic test bed for investigating automatic and voluntary processes in random sequence generation. *Frontiers in Psychology 11*, 551126.

Dyson, B. J. (2019). Behavioural isomorphism, cognitive economy and recursive thought in non-transitive game strategy. *Games 10*(3), 32.

Elo, A. (1978). *The rating of chess players, past and present*. New York: Arco.

Gobet, F., de Voogt, A. J., & Retschitzki, J. (2004). *Moves in mind: The psychology of board games*. Hove, UK: Psychology Press.

McHale, I. and A. Morton (2011). A Bradley-Terry type model for forecasting tennis match results. *International Journal of Forecasting 27*(2), 619–630.

N'Guessan A., G. (1992). *Mécanismes d'apprentissage de l'awèlé.* Fribourg: Editions universitaires.

North, B. V., D. Curtis, and P. C. Sham (2002). A note on the calculation of empirical p values from Monte Carlo procedures. *The American Journal of Human Genetics 71*(2), 439–441.

Retschitzki, J. (1990). *Stratégies des joueurs d'awèlé.* Paris: L'Harmattan.

Roberts, J., M. Arth, and R. Bush (1959). Games in culture. *American Anthropologist 61*(4), 597–605.

Schauberger, G., A. Groll, and G. Tutz (2018). Analysis of the importance of on-field covariates in the German Bundesliga. *Journal of Applied Statistics 45*(9), 1561–1578.

Serra, L. (2020). Preservation of Mediterranean intangible cultural heritage through virtual gaming and informatics: The case of Sardinian Mùrra. *Systemics, Cybernetics and Informatics 18*(6), 24–30.

Towse, J. N. and D. Neil (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods, Instruments, & Computers 30*(4), 583–591.

Turner, H. and D. Firth (2012). Bradley-Terry models in r: the bradleyterry2 package. *Journal of Statistical Software 48*(9).

Yang, T. Y. (2004). Bayesian binary segmentation procedure for detecting streakiness in sports. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 167*(4), 627–637.

Zhang, H., F. Moisan, and C. Gonzalez (2021). Rock-paper-scissors play: Beyond the win-stay/lose-change strategy. *Games 12*(3), 52.

Received: 19 August 2023
Revision received: 23 November 2023
Accepted: 3 December 2023

Appendix: Derivation of B_{10}

For independent binomial data $D \equiv y_1, \dots, y_n$ with parameters $(m_i, p_i), i = 1, \dots, n$, we want to estimate whether or not there exists a cut point $c \in \{1, \dots, (n-1)\}$, such that the event probabilities $p_i = p_1$ for $i \le c$ and $p_i = p_2$ for $i > c$. To test the null model $M_0: p_1 = p_2$ against the one-cut-point model $M_1: p_1 \ne p_2$, we calculate the posterior ratio $\frac{P(M_1|D)}{P(M_0|D)} = \frac{P(D|M_1)\cdot P(M_1)}{P(D|M_0)\cdot P(M_0)}$. Under the assumption that the two models have the same prior probability, the selection criterion simplifies to $\frac{P(D|M_1)}{P(D|M_0)}$, the Bayes factor of model M1 against M0, denoted by B10.

We next explain how to calculate P(D|M1) and P(D|M0) for our data (a more general derivation can be found in (Yang, 2004)). In our application, data consist of binary outcomes (i.e., win or lose, left or right hand). The likelihood under model 1 is given by

$$L_1(c, p_1, p_2, M_1) \equiv P(D|c, p_1, p_2, M_1) = \prod_{i=1}^{c} p_1^{y_i}(1-p_1)^{1-y_i} \cdot \prod_{i=c+1}^{n} p_2^{y_i}(1-p_2)^{1-y_i},$$

and the likelihood under model 0 (assuming p1 = p2 = p0) is given by

$$L_0(p_0, M_0) \equiv P(D|p_0, M_0) = \prod_{i=1}^{n} p_0^{y_i}(1-p_0)^{1-y_i}.$$

We assign independent uniform (0,1) priors for the parameters p_0, p_1, p_2 and an independent discrete uniform prior for c (in general, beta priors can be used for p_0, p_1, p_2). We then have

$$\begin{aligned}
P(D|M_1) &= \sum_{c=1}^{n-1} \int_0^1 \int_0^1 P(D, c, p_1, p_2|M_1)\ d_{p_1} d_{p_2} \\
&= \sum_{c=1}^{n-1} \int_0^1 \int_0^1 P(D|c, p_1, p_2, M_1) \cdot P(c|M_1) \cdot P(p_1|M_1) \cdot P(p_2|M_1)\ d_{p_1} d_{p_2} \\
&= \sum_{c=1}^{n-1} \int_0^1 \int_0^1 L_1(c, p_1, p_2, M_1) \cdot \frac{1}{n-1}\ d_{p_1} d_{p_2} \\
&= \frac{1}{n-1} \sum_{c=1}^{n-1} \int_0^1 p_1^{\sum_{i=1}^{c} y_i}(1-p_1)^{c-\sum_{i=1}^{c} y_i}\ d_{p_1} \cdot \int_0^1 p_2^{\sum_{i=c+1}^{n} y_i}(1-p_2)^{(n-c)-\sum_{i=c+1}^{n} y_i}\ d_{p_2} \\
&= \frac{1}{n-1} \sum_{c=1}^{n-1} \frac{(\sum_{i=1}^{c} y_i)!\,(c-\sum_{i=1}^{c} y_i)!}{(c+1)!} \cdot \frac{(\sum_{i=c+1}^{n} y_i)!\,((n-c)-\sum_{i=c+1}^{n} y_i)!}{(n-c+1)!}.
\end{aligned}$$

The last step is calculated by matching the integrals to beta kernels. Similarly, for P(D|M0), we have

$$P(D|M_0) = \int_0^1 P(D, p_0|M_0)\ d_{p_0} = \int_0^1 p_0^{\sum_{i=1}^{n} y_i}(1-p_0)^{n-\sum_{i=1}^{n} y_i}\ d_{p_0} = \frac{(\sum_{i=1}^{n} y_i)!(n-\sum_{i=1}^{n} y_i)!}{(n+1)!}.$$

If model M1 is selected, we want to estimate the cut point c. The posterior distribution of c is given by

$$\begin{aligned}
P(c|D, M_1) &\propto \int_0^1 \int_0^1 P(D|M_1, c, p_1, p_2) P(M_1|c, p_1, p_2) P(c) P(p_1) P(p_2)\ d_{p_1} d_{p_2} \\
&\propto \int_0^1 \int_0^1 L_1(c, p_1, p_2, M_1)\ d_{p_1} d_{p_2} \\
&= \frac{(\sum_{i=1}^{c} y_i)!\,(c-\sum_{i=1}^{c} y_i)!}{(c+1)!} \times \frac{(\sum_{i=c+1}^{n} y_i)!\,((n-c)-\sum_{i=c+1}^{n} y_i)!}{(n-c+1)!},
\end{aligned}$$

$c = 1, \dots, n-1$. We can use the posterior mean to estimate the cut point c, i.e.,

$$\hat{c} = \frac{\sum_{c=1}^{n-1} P(c|D, M_1) \cdot c}{\sum_{c=1}^{n-1} P(c|D, M_1)}.$$

It is possible that more than one cut points exist in the data. An iterative segmentation process can be used to detect multiple cut points (Yang, 2004). In our application, however, we consider only the one-cut point model due to the relatively small sample sizes.