

MEASURING DRAMA IN GOOSE-LIKE GAMES

J.P. Neto & J.N. Silva*

Abstract

For games of complete information with no chance component, like Chess, Go, Hex, and Konane, some parameters have been identified that help us understand what makes a game pleasant to play. One of these goes by the name of *drama*.

Briefly, drama is linked to the possibility of recovering from a seemingly weaker position, if the player is strong enough. This is an important requirement to prevent initial advantages to be amplified into unavoidable and thus uninteresting victories. Drama is a feature that arguably good board games should have, since it is relevant in the perception of the play experience as pleasant.

Despite its intrinsic qualitative nature, we suggest the adaptation of the concept of drama to games of pure chance and propose a set of objective criteria to measure it. Some parameters are here used to compare Goose-like games, which we compute via computer simulation for some well-know games. A statistical analysis is performed based on the play of millions of matches done by computer simulation. The article discusses correlations and patterns found among the collected data. The methodology presented herein is general and can be used to compare other types of board games.

Drama and other qualities of abstract games

Drama is a subjective measure concerning the quality of a board game. Mark Thompson in his seminal article defined *drama* in the following way (Thomson (1999)):

[...] it should be possible for a player to recover from a weaker position and still win the game. Victory should not be achievable in a single successful blow; the suspense should continue through an extended campaign. Otherwise an early disadvantage makes the remainder of the game uninteresting: the doomed

*The authors thank the referee for some very important corrections and suggestions for further work.

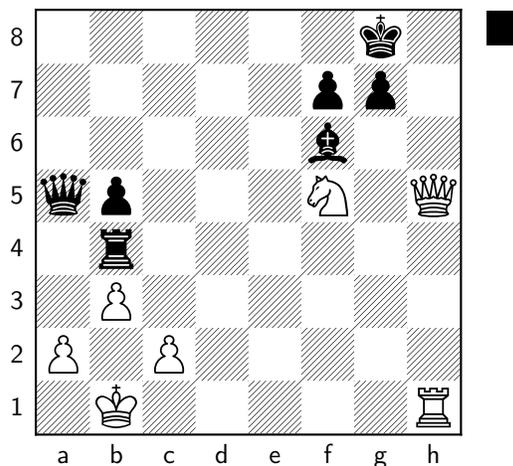
player rightly guesses that the puzzle he is trying to solve has no solution and that thinking about it is futile. A game's drama might be measured roughly by matching a strong player against a weak player, and having them switch sides after the strong player achieves an advantage. In a dramatic game the strong player will still have a chance of winning.

Thompson's text focused mainly on perfect-information games like Chess, as shown in the following section. However, we consider that the concept is also relevant in games where the random factor replaces the player's skill.

Chess is dramatic

Alexander Aleksandrovich Alekhine was a great chess player who lived a very intense life during eventful times. Born in Russia in 1892, he died in Portugal in 1946 under mysterious circumstances. Alexander Alekhine was world champion from 1927–35 and 1937–46. We will use one of his games, against a weaker opponent, to illustrate our point.

Their game reached the following position, where Alekhine conducts the white pieces.



Alekhine's adversary, facing the threat of mate in **h8** considered

1... g6 2 ♕h7+ ♖f8 3 ♕h8+ ♗h8 4 ♜h8# and resigned.

Alekhine suggested they switched colors, and played, **1... ♜h4**

Now conducting the white pieces, the player did not find any nice way of capturing the black rook:

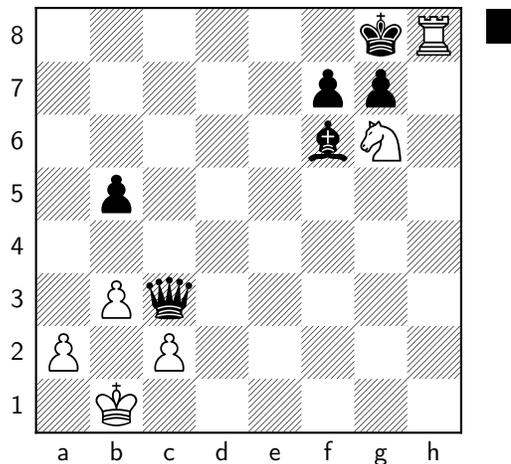
2 ♖h4 ♙h4+

2 ♜h4 ♚e1+

2 ♘h4 ♗c3 3 ♙c1 ♖a1+ 4 ♘d2 ♗h1+

therefore, he resigned. But Alekhine suggested a new swap! Playing White again, the world champion played 2 ♘h4 and the game unfold 2... ♗c3 3 ♗h8+! ♙h8 4 ♘g6+ ♙g8 5 ♜h8#.

The final position:



Two swaps such as these could not happen in many games. However, Chess is really very dramatic. No wonder newspapers carry Chess problems on a regular basis.

Race games

Race games have a long and rich history. One of the oldest is the mysterious Mehen (Parlett, 1999, p. 89), from ancient Egypt. Snakes & Ladders (Parlett, 1999, p. 91), of Eastern origin, is a race in which the winner is the player that lands on positively charged cells of the board, morally speaking.

The Gioco dell'Oca was a board game offered by Francesco de Medici, grand-duke of Tuscany from 1574 to 1587, to the king of Spain, Filipe II. As Parlett tells us (Parlett, 1999, p. 95), "Goose was little more than a lottery, being entirely governed by chance and invoking not the slightest element of skill or true player interaction towards the winning of the stakes."

Surprisingly, its details and procedure remain constant to this day. The survival of this game is very interesting, as its mechanics seem to be based not only on ludic principles, but also on the medieval cabala, according to Adrian Seville, who has been studying Goose games for a long time (Seville (2009), Seville (2016)).

The authors noticed, namely among students of the University of Lisbon, that Goose is very enjoyable. How can this be, if the game is completely automatic and random?

There are more people having fun playing the game of Goose than betting on coin tossing. However, chance rules both games completely.

Those of us that played Goose know that the geese, the Inn, the Bridge, among others, are squares that influence our fate in the game. . . we want to land in some of these and avoid others. The cells of the board come with emotions attached.

The accidents that characterize these games, like snakes, ladders, geese, etc. add emotional complexity to the games. Falling behind or getting ahead are not absolute states, fate can play a part and turn the wheel of fortune. Of course, a pleasant game should be rich in such turns of events, but not too much so. To be behind must keep some of its meaning, as must being ahead of other players, otherwise each game is an exercise in chaos, too messy to be played emotionally.

The strong random element in race games, instead of being an extra difficulty for analysis, provides an opportunity to measure drama using probabilistic tools. Instead of focusing on criteria that would provide a concrete number summarizing some pattern from the game rule structure, we will address concepts like averages and standard deviations that will naturally arise from playing many matches. This allows us to replace mathematical analysis with computational simulations.

Taking this into account, the authors reached the following statistical criteria for drama:

- How balanced are the player's winning percentages? Given n players, let's define WIN as the standard deviation of the estimated winning percentages of the first, second, third, . . . , players. If they are all equal, then WIN will be zero. If there are large differences in these estimates (e.g., the first player has a great advantage by starting), then WIN will increase. A more balanced game allows for more dramatic matches, since the element of surprise of who will win increases with it.
- LEAD is the number of turns on which the winner had the lead just before calling the game. If LEAD is, on average, very high, it means

that the last turns on a race game lack emotion and tension. The leader just keeps leading until the game is over. On the other hand, if LEAD is low, last turn surprises will be common, adding to the suspense of the game.

- The average number of players that take the lead during a match at the end of an entire turn (i.e., after all players' moves) is another parameter — ANP. This applies better to games with a reasonable number of players (say, at least five). Naturally, a low ANP means very few players were on the lead, which is bad for dramatic matches. A very high ANP might not be good; if, on average, almost all players take the lead, that means the game is not very decisive, and there is little tension in the initial stages, since players know that, on average, all will be ahead at some point. This would also imply very long matches (as always, on average). It seems appropriate that, for n players, ANP should be a middle value between 1 and n .

Adrian Seville had already considered the use of Monte Carlo methods in Seville (2002) to count the number of rounds to win a Game of Goose and a game of Snakes and Ladders.

The Goose Experience

The experimental setup

Here we consider race games of the Game of Goose family. The game of Goose has special squares (showing geese) on numbers 5, 9, 14, 18, 23, 27, 32, 36, 41, 45, 50, 54, and 59. If a piece lands on a goose, it moves by the same amount again (forward or backward). If this causes the piece to land on another goose, it moves again in the same way. Also

- The course contains 63 cells;
- Each player has one token that moves accordingly to the sum of the throw of two dice;
- Each space in the board can only be occupied by a single token;
- If a token lands on a previously occupied space, the two tokens switch their positions;
- A token can only end the game if, in the last turn, the dice sum is the exact number needed to reach the final space. If the number exceeds the final space, there is a rebound effect and the remaining moves are made backwards;

- The accidents, besides the geese, are at cells: 6 (The Bridge) - Go to space 12; 19 (The Inn) - Miss one turn; 31 (The Well) - Wait until someone comes to pull you out, they then take your place; 42 (The Maze) - Go to space 39; 52 (The Prison) - Wait until someone comes to release you, they then take your place; 58 (Death) - Go off the board, start the game again.

The games tested were the Game of Goose and two of its variants of pedagogical nature: Le Jeu de la Sphère ou de l'Univers selon Tyco-Brahe¹ and The New Navy Game.² The Navy game is very like Goose, with similar traps but only one series of goose-like spaces, whereas the Universe game has a completely different set of traps and no regular series of favorable spaces. Both, however, do have the same rule as Goose for overthrows of the winning space. We will denote these three games by *historical variants*. We also included an *abstract variant*, that uses a board of 63 spaces like the games of Goose and Navy, without any special rules (no traps, no jump spaces, etc.). This abstract variant is useful as a kind of control to check the impact of the special rules that characterize every historical variant.

A program was implemented, written in the statistical R language, to automatically play these Goose-like games.³

We performed multiple trials for each game. A *trial* consists of 40,000 matches played by the computer. We performed 12 trials for each one of the historical variants, and 6 trials for the abstract variant for each number of players, from three to eight. The number of total matches played was over 10 million. This large amount was intended to decrease the probability of finding spurious patterns due to random noise effects.

For each trial a number of statistics were gathered for posterior analysis in order to estimate values for LEAD, WIN, and ANP drama criteria. The next sections summarize these results.

WIN criterion

The WIN criterion deals with how balanced the variants are, concerning winner percentages for each player. The plot at Figure 1 shows the results obtained by the match simulations performed by our software.

All historical games show a decrease in variation as the number of players increase. This means we find higher levels of drama with more players

¹<http://www.giochidelloca.it/scheda.php?id=1448>

²<http://www.giochidelloca.it/scheda.php?id=1204>

³The source code can be accessed at <https://github.com/jpneto>

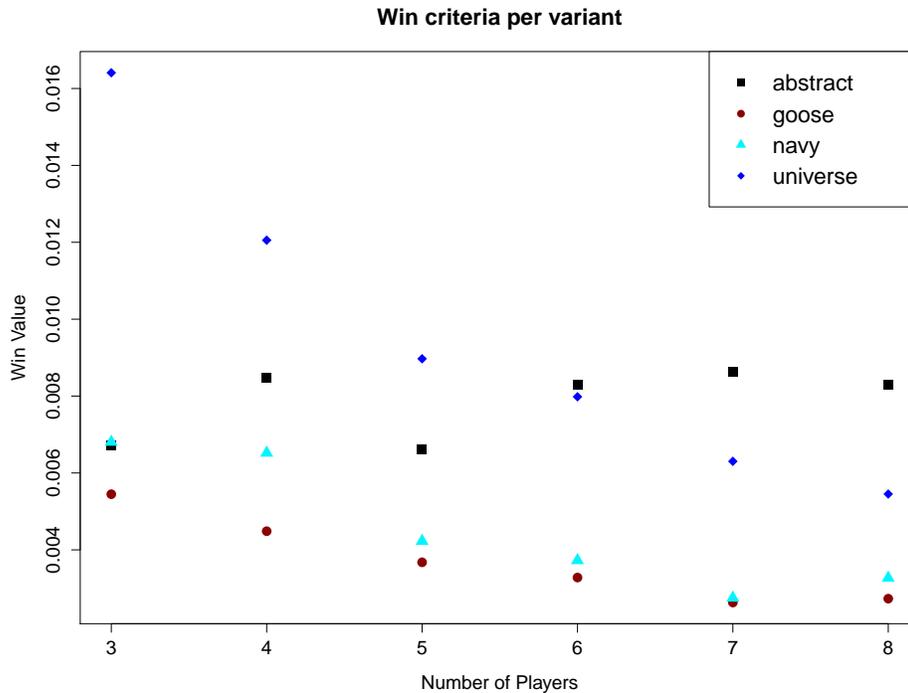


Figure 1: This plot shows the values of WIN for each variant — each identified by a specific color and shape — and for each number of players. The WIN values are the standard deviation of the mean winner percentages of the respective game trials.

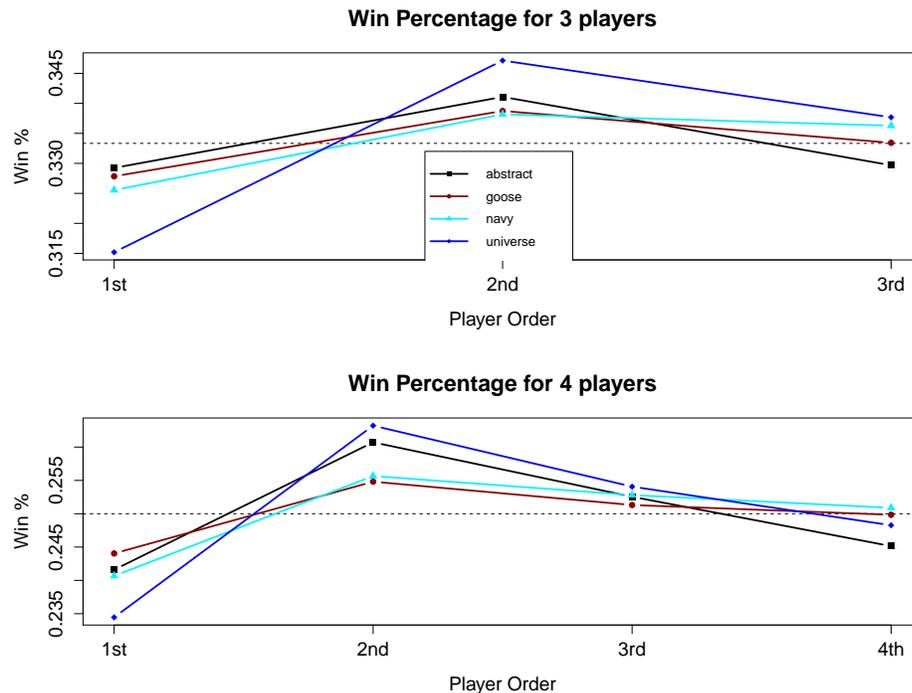
according to WIN criterion. The same effect does not happen in the abstract version. This indicates that the special rules typical of the historical variants do play a relevant role in adding drama to the overall gaming experience.

The game of Goose achieves the best WIN values for every number of players. The game of Navy achieves similar results. The other two games produce a larger variation. Notice, however, that we only got small WIN values, almost all between 0.004 and 0.009. There are only two outliers for the Game of Universe with three and four players. These are values quite close to zero. To provide some context, a game with four players where the first two players had chances of victory of 26% and the third and fourth players had 24%, the WIN value would be 0.012.

There appears to be a curious inflection point at seven players. Adding an eighth player seems to increase the variation, if only slightly, for the games of Goose and Navy. This might be a consequence of the board size.

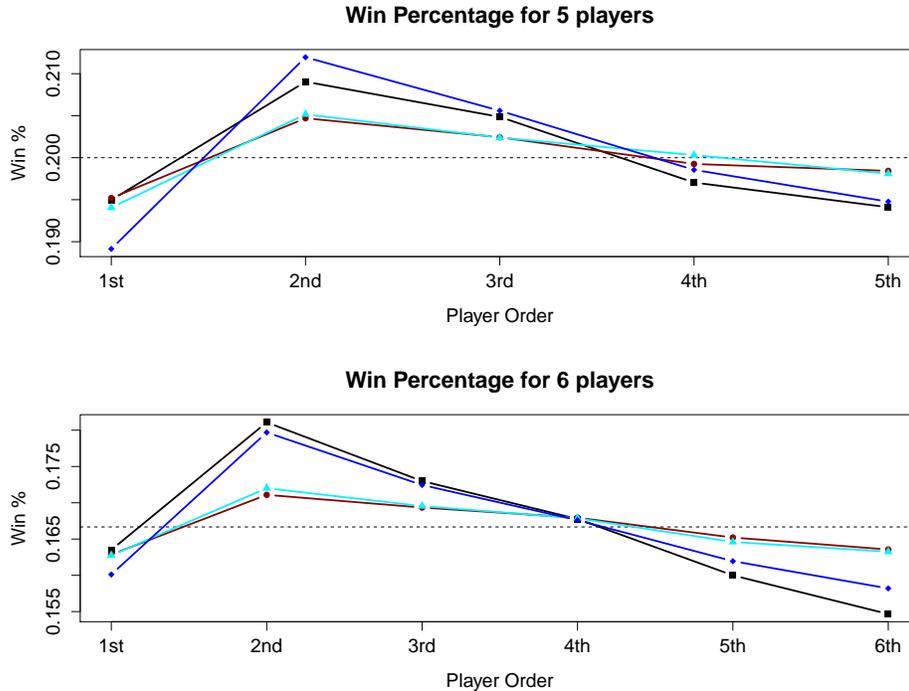
Both games have 63 spaces and possibly the game becomes too crowded. That inflection does not happen for the game of Universe, which has a larger board of 70 spaces. We speculate that its inflection point — if indeed, this inflection is not just a product of statistical noise — will occur by adding a ninth or tenth player.

Higher WIN values means that some players are winning more than others. Figure 2 shows which players are getting more wins.



Looking at these plots, the most striking feature is that the second player has a constant advantage, regardless of the number of players or which game variation is being played. The third player also has some advantage but it is milder. Also surprising is the disadvantage of the first player, who only achieves mean performance with seven and eighth players. Perhaps not so surprising is the disadvantage of the last players, especially in game with six players or more, given that when they start, there are already several tokens on the board.

These plots confirm the better results of the Game of Goose, since its lines are closer to the theoretical best result — the dotted black line where all players would lie if all had the same winning percentage.



Also, the worst results from the game of Universe and the abstract variant can be seen here too, since their slopes are steeper, giving larger advantages for the second and third players — and thus making these two variants less fair to play.

To confirm that the second player's advantage is not a product of random noise, we tested these numbers against an uniform distribution of values — which would be the appropriate statistical model for a perfect balanced race game, let us denote it by *balanced variant* —, to see if the differences between the theoretical value and the simulated one were indeed significant. For each pair variant/number of players, we made a trial of 40,000 matches of this *balanced variant* and computed the ratio between the player with more wins against the player with less wins. These ratios simulate the outcomes of a perfect fair game.⁴ We repeated this procedure 2000 times to reduce variation. Then, given these values, we found at which percentile our results fitted among these 2000 balanced outcomes.

⁴Notice that we did not play a game here. For n players, we just sampled numbers from a uniform distribution between 1 and n and assigned the outcome i as a win for the i^{th} player

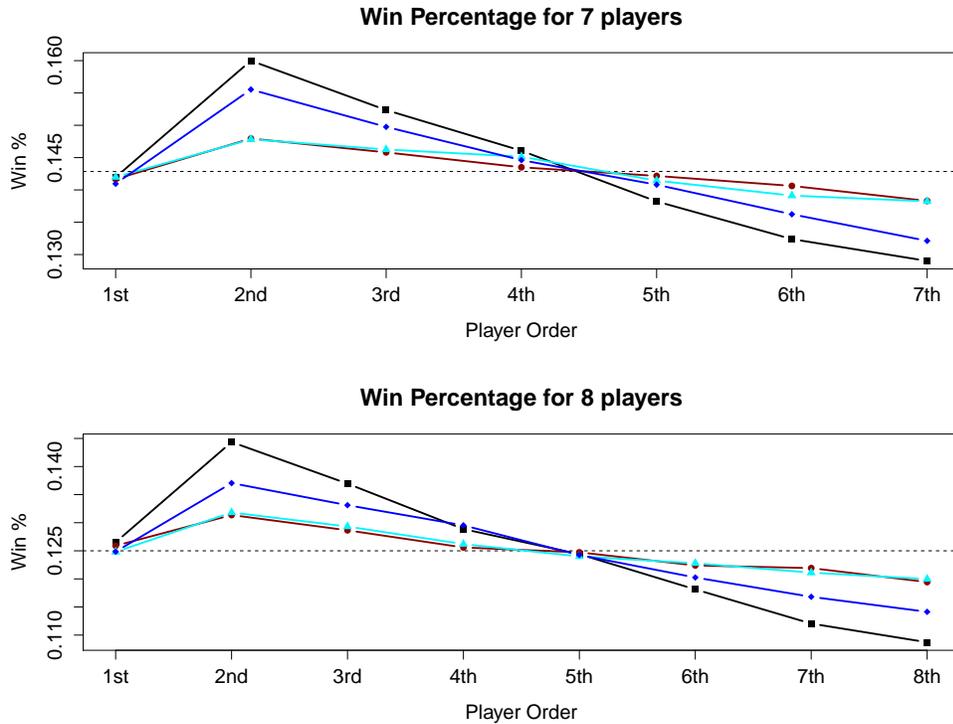


Figure 2: Each plot shows the winning percentages of all four simulated variants for a given number of players. The dotted horizontal line marks what would be a perfect balanced game. This means that curves closer to that theoretical line are more balanced than others.

If we get a value near percentile 100 it means that our result is non uniform with high significance (i.e., some player has advantage over others). We put the threshold at percentile 90. Less than this value is interpreted as not providing enough evidence to reject the hypothesis that a given variant is a fair game.

Figure 3 shows the computed percentiles: There are 36 trials for the abstract variant, and 72 trials for each of the historical games. The values below the dotted line could be interpreted as evidence for the uniform hypothesis (except for the game of Universe that does not have a single outlier). However, the great majority of the results were placed at percentile 100 or at very high 90s, which is a strong evidence for the non-uniformity hypothesis. These data, together with the previous plots, presents a convincing argument for the second player advantage in these four variants.

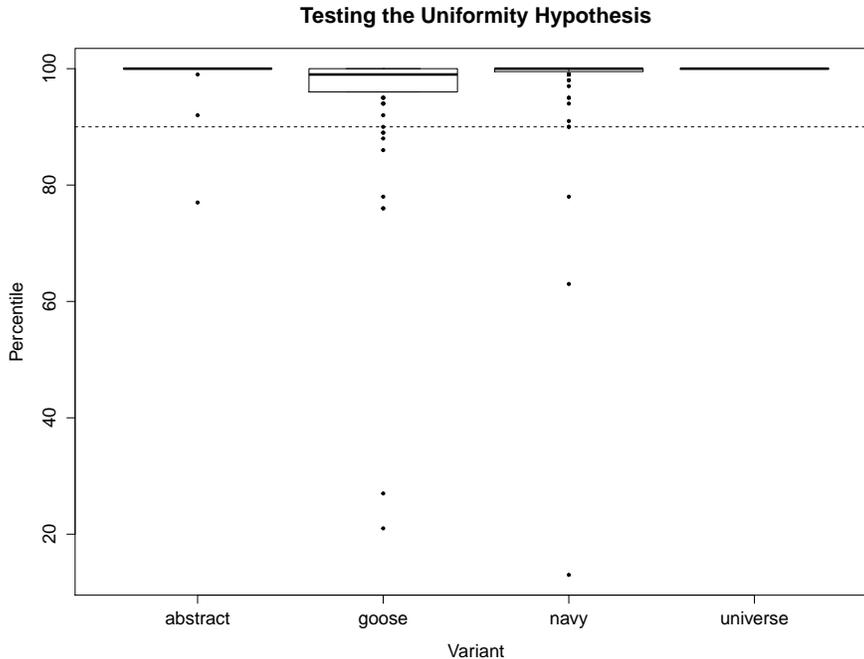


Figure 3: These are statistical box plots. The box represents the inter-quartile range, IQR, i.e., the values from percentile 25 to 75 (the bold line represents the sample median). Notice that for all but the Game of Goose, the results produced a IQR so small that it collapsed the box into a line. The points represent values outside IQR. The dotted line marks the chosen threshold at percentile 90.

However, even if these variants show that some players have initial advantages, the difference between the estimated best and worst chances of winning is not that large. The maximum differences are around three percentage points, which is not enough to impact on the game play of occasional players. Seasoned players might notice this, but can easily fix this problem by changing the order of playing on a regular basis.

LEAD criterion

LEAD values state that a game is more dramatic if the player that wins did achieve the lead only a few turns before the end of the game. In the following simulations we measured the lead up to five turns before the end of the variant. Again, we did this for each variant and each number of players from three to eight.

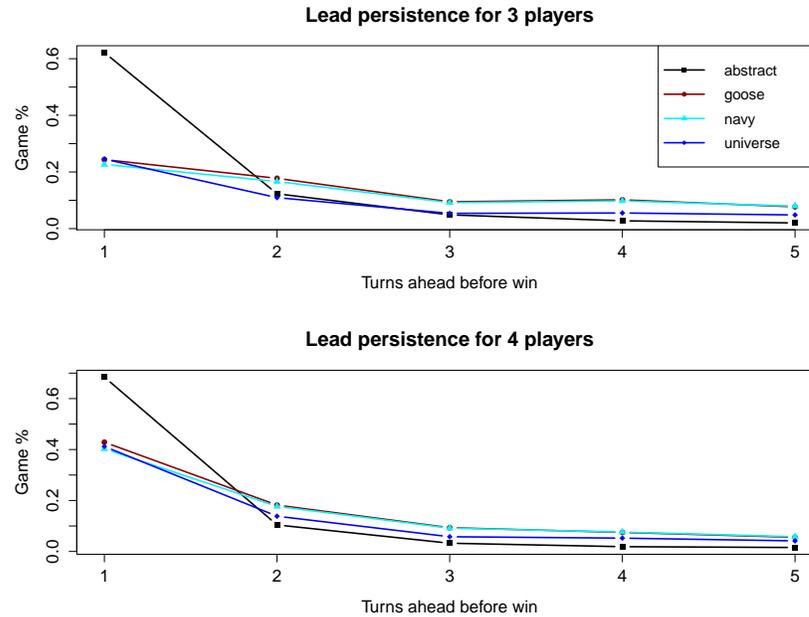
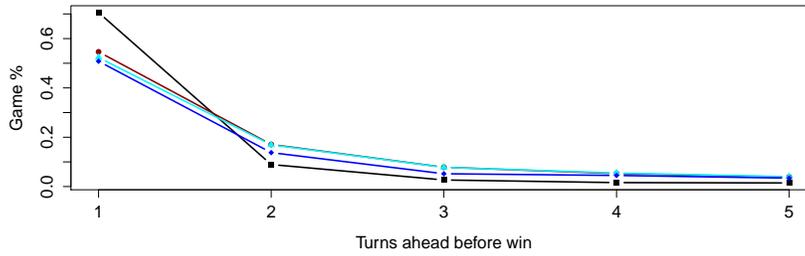


Figure 4: Each plot shows what the match percentage where the winner had the lead for one to five turns at the end. As an example for 3 players, the abstract variant, in 61% of the matches, found its winner only at the last turn, i.e., the majority of matches had seen a leader reversal just at the last moment. This implies good drama from a LEAD perspective.

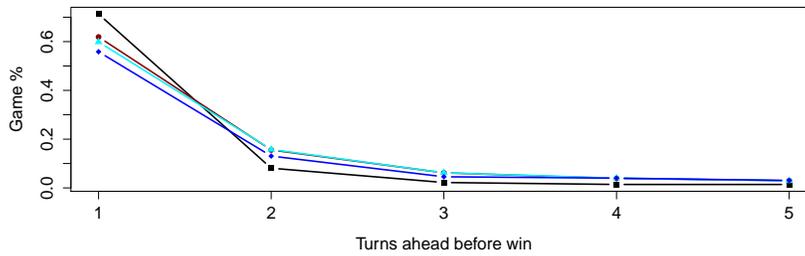
The abstract variant is the most dramatic by this criterion. Even for three players, the games where the winner just got his lead in the last turn are more than 60%. However, this difference gets smaller as we add more players. With six players the distributions for all four games are quite similar. Comparing the three historical games, the Game of Goose is slightly better than the other two, but this difference might not be significant despite the number of matches we are dealing with.

Among all games, the abstract version had 84% of games with up to five turns ahead for the winner, the Goose had 86%, the Navy 84% and the Universe 74%. This means all variants are quite dramatic by this criterion. They do not frequently produce matches where the winner is known many turns before the end. Only on one sixth to one fourth of the games the lead is kept for more than five turns. Curiously, if these percentages were more extreme, say, if 95% of matches were won at the last moment, the drama would also be hard to keep: players would rightly guess that the current

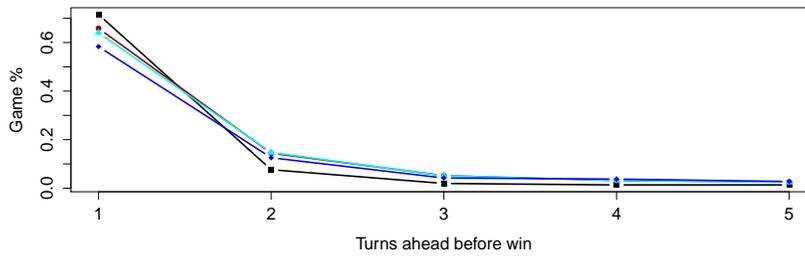
Lead persistence for 5 players



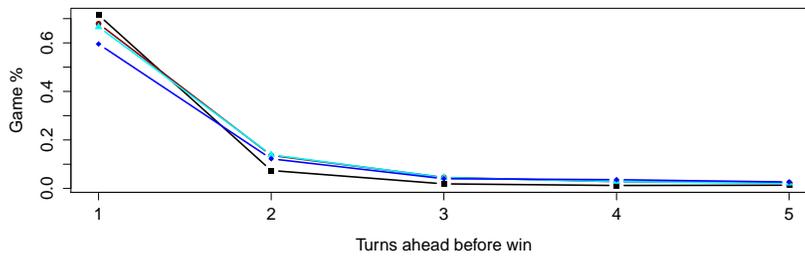
Lead persistence for 6 players



Lead persistence for 7 players



Lead persistence for 8 players



lead would not win. As it is, the players still have plenty of room for doubt, which is a good feature for a race game.

ANP criterion

The ANP criterion focus on the number of different players that were ahead during a game. The argument is that a game is more dramatic the more players are able to take the lead. A game where the first player leading maintains her lead during the entire match is arguably less dramatic than one where there are many players able to be ahead of the others during the match. The mean values of all trials are presented at Figure 5.

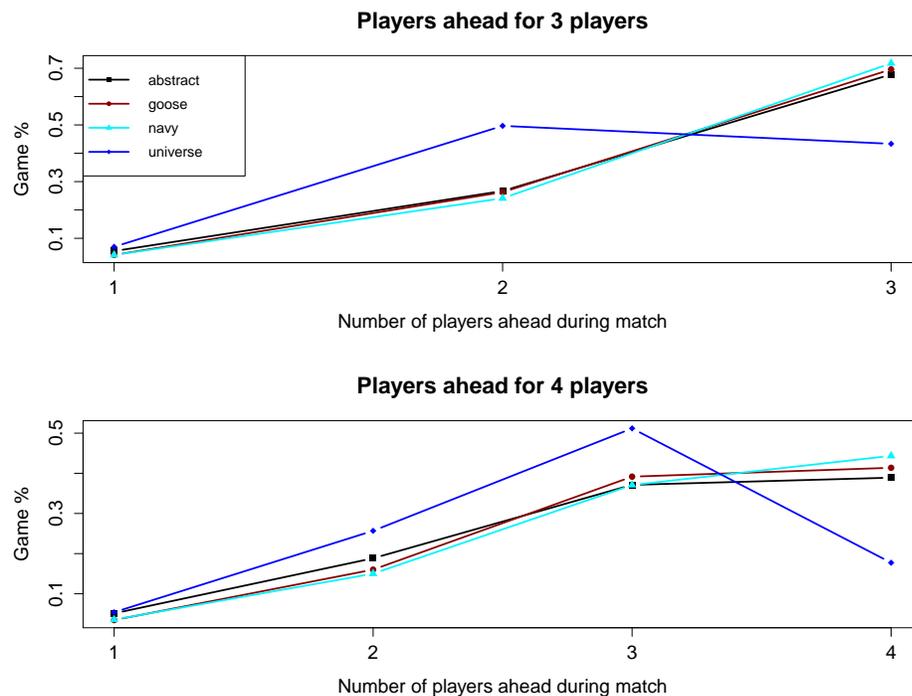
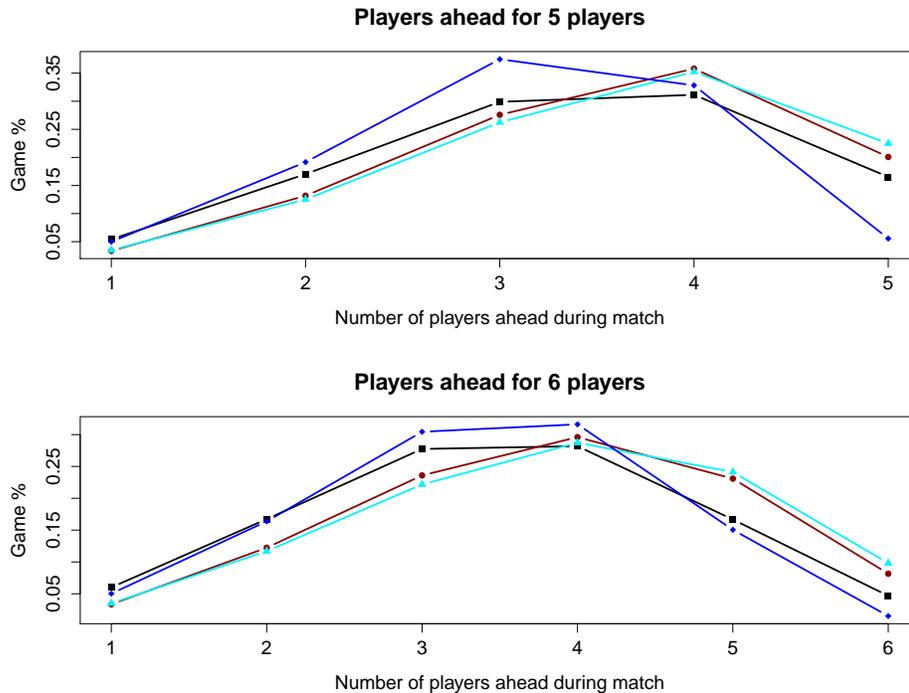


Figure 5: Each plot shows, for each game and for n players (again, $3 \leq n \leq 8$), the percentage of games where the number of leaders were 1 to n . As an example, the Game of Universe with 3 players had around 50% of the matches with two different leaders.

For all variants, the proportion of matches with a single lead is quite small — fewer than 5% of the games are like that. The data shows that in the typical match, with at least four players, about half the players take the



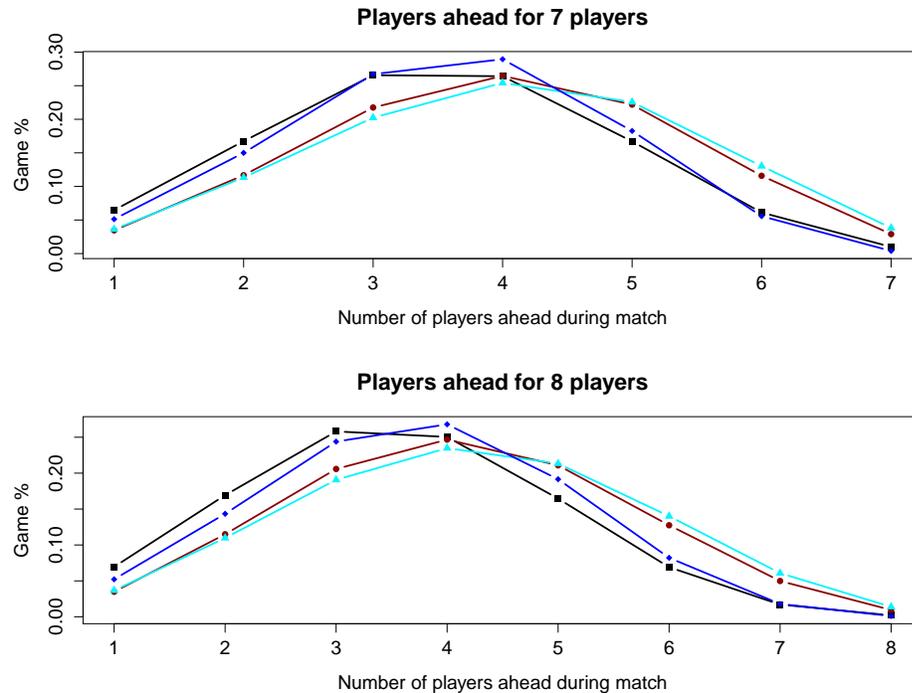
lead. This is a good balance. A game where, on an average match, all the players are ahead at some point might feel a bit too chaotic (and indeed, there are very few of those). But a game where about half the players take the lead provides, in our perspective, a good gaming experience.

Comparing games, the game of Universe is the less dramatic for three or four players, but behaves quite similarly to the others if the number of players is five or more. For many players, the Game of Goose and the Game of Navy are the ones that allow more lead changing events.

Game Length

This experience provided us with more data besides the evaluation of the three drama criteria. Another feature, not related to drama, but important in gaming experience, is the average match length. If too short or too long matches are frequent, the game becomes boring and dull. However, if the average length is usually socially admissible,⁵ the fact of having some rare

⁵The range of admissible turns a game should have depends on the culture embedding it. It is quite possible that if Go was invented today it would not have succeeded given the



matches ending too soon or taking too long might be interpreted as a game bonus and played with increased interest.

Figure 6 shows an histogram of the distribution of game moves for 12,500 games of four player Goose. The histogram has the shape of an asymmetric distribution with an heavy tail to the right: some games take many hundreds of moves to end. Disregarding this tail, i.e., rare matches that take too long due to a great number of consecutive traps and captures, the histogram shape is quite close to a log-normal distribution. Figure 7 shows the best fit for a log-normal. This fit also works nicely to all the historical variants as well to the abstract variant.

This relationship with the log-normal is just an empirical statement. We do not know a formal argument relating this well-known distribution to the number of moves necessary to end a Goose-like race game. It is quite possible that more Goose variants also share this pattern, and it should be an interesting subject to determine if this pattern also arises in other race games.

great amount of turns needed to complete a match (around 200 to 300 stone placements).

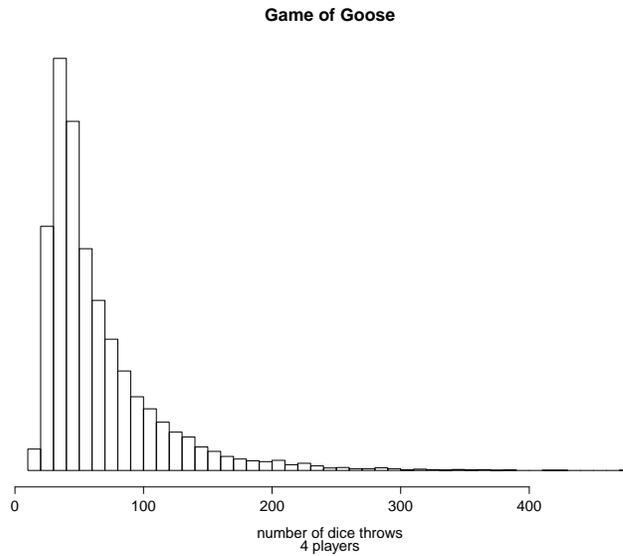


Figure 6: This plot shows the histogram of the number of needed dice throws to end 12500 games of Goose with four players. Herein, a turn consists of four dice throws, since there are four players.

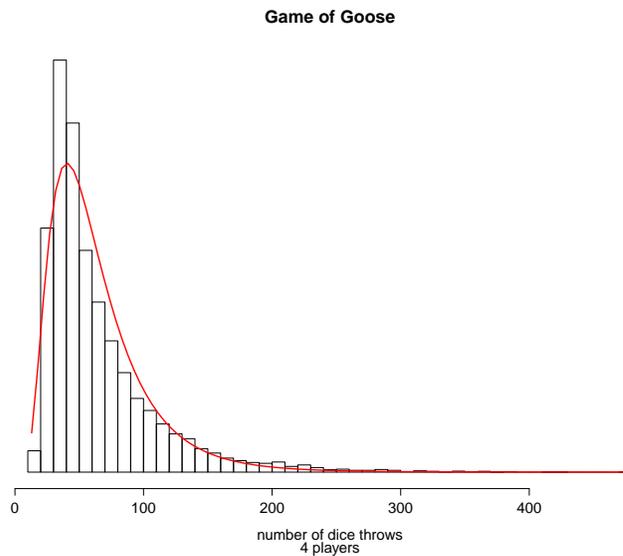


Figure 7: This plot shows the previous histogram with the best log-normal fit.

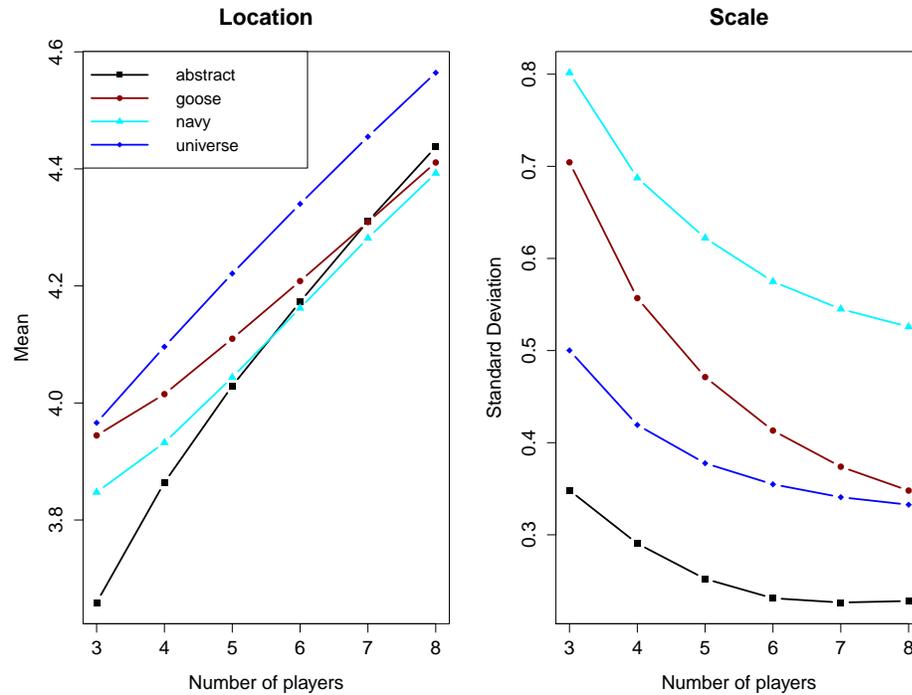


Figure 8: A log-normal distribution has two parameters: i) a location parameter that specifies the distribution’s mean, and ii) a scale parameter that specifies how much spread we should expect. This pair of plots show, for each variant, how these parameters change accordingly to the number of players.

According to Figure 8, all variants despite having quite different parameters regarding their special rulings (e.g., where are the traps, where and how are the jump spaces), share a similar behavior concerning the log-normal fits. The location trend suggests that Goose-like variants with more players tend to take longer to finish, which is a natural result to expect. The scale trend, however, suggests that the expected variation of the number of turns tends to be smaller as the number of players increase, which is not an obvious consequence, even if it seems to stabilize after a certain amount of players (its exact number depends on the variant). Both trends are well-behaved. This is evidence that it should be possible to achieve good estimations for the expected number of moves, given the number of players.

Final remarks

This paper presented a methodology that focus on computational simulation to find patterns in the dynamics of race games. This methodology requires the implementation of the chosen set of games and the definition of a set of criteria to be measured.

Here, the chosen games were all related to the Goose family of games, but any other race game – and variants – can be selected. This methodology is general enough to support any stochastic game dynamics. The chosen criteria were designed to be quantitative measurements of game’s drama.

We also measured the length of games and found an unexpected relationship with the log-normal distribution, which we cannot explain analytically but the empirical evidence is very strong. This fit works nicely for the historical variants as well as for the abstract variant, suggesting that it arises from their main common feature: the ‘rebound’ or ‘reverse-overthrow’ rule for dealing with excess points at the end of the track. This is a topic for future work.

References

- Ciampi, L. & Seville, A. (n.d.). *Giochi dell’Oca e di percorso*. Retrieved from <http://www.giochidelloca.it/index.php>
- Parlett, D. (1999). *The Oxford History of Board Games*. Oxford University Press.
- Seville, A. (2002). *Statistical characteristics of enjoyable race games*. Barcelona, BGS Colloquium. (Revised in 2011. Available at Ciampi, L. & Seville, A. (n.d.))
- Seville, A. (2009). The Sociable Game of the Goose. In Silva, J.N. (Ed.), *Proceedings of the Board Game Studies Colloquium XI* (pp. 3–17).
- Seville, A. (2016). *The Royal Game of the Goose: four hundred years of printed board games*. The Grolier Club.
- Thomson, J. M. (1999, July). Defining the Abstract. *The Games Journal*. Retrieved from <http://www.thegamesjournal.com/articles/>

Received on January 22, 2016, accepted on September 10, 2016.