An Analysis of Majority Voting in Homogeneous Groups for Checkers: Understanding Group Performance through Unbalance

Danilo S. Carvalho[⋆], Minh Le Nguyen, and Hiroyuki Iida

Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi City,
Ishikawa, Japan,
`{danilo, nguyenml, iida}@jaist.ac.jp`

Abstract. Experimental evidence and theoretical advances over the years have created an academic consensus regarding majority voting systems, that the group performs better than its components, under certain conditions. However, the underlying reason for such conditions, e.g. stochastic independence of agents, is not often explored and may help to improve performance in known setups by changing agent behavior, or find new ways of combining agents to take better advantage of their characteristics. In this work, an investigation is conducted for homogeneous groups of independent agents playing the game of Checkers. The analysis aims to find the relationship between the change in performance caused by majority voting, the group size, and the underlying decision process of each agent, which is mapped to its source of non-determinism. A characteristic unbalance in Checkers, due to an apparent initiative disadvantage, serves as a pivot for the study, from which decisions can be separated into beneficial or detrimental biases. Experimental results indicate that performance changes caused by majority voting may be beneficial or not, and are linked to the game properties and player skill. Additionally, a way of improving agent performance by manipulating its non-determinism source is briefly explored.

Keywords: Group Performance; Checkers; Homogeneous Groups; Majority Voting

1 Introduction

Studies on the performance of groups of independent agents on select activities have developments as early as late 19th century [12]. Such works have addressed various aspects of group activity phenomena, such as independence, collaboration, information sharing [2], common knowledge construction [14], among others. To conduct such investigations, data regarding different characteristics of a group and its participants are collected and then interpreted in the light of the observed facts. For human groups, age, gender and estimated skill in the

---

[⋆] ABC

task of interest are examples of properties that can be observed and related to group behavior [14, 13], as well as group variables, such as size, age and frequency of interaction. This often results in well-founded postulates about certain group phenomenon in set conditions, but lacks in explaining the reasons of such phenomena. One such case is the consensus regarding majority voting systems, that the group performs better than its components, when those are statistically independent. Understanding the cause of this behavior would enable better prediction capabilities, given enough prior information about the group, agents and the task of interest. It would also allow steering behavior more effectively when the group properties can be manipulated, e.g., changing group members. This point of view is, however, not often explored in game research literature, due to the complexity of group interactions, specially involving humans.

In this work, an investigation is conducted on the underlying factors of homogeneous group performance, using a computer game of *Checkers* as the task of interest. The choice for the Checkers computer game is threefold: i) it is simple enough to perform extensive analysis of game properties and results, while being complex enough to have a wide array of possibilities and still expensive to search exhaustively, despite being solved computationally [8]; ii) it allows isolation of group and agent properties, so that statistical analysis is facilitated; and iii) it allows manipulation of agents' properties defining their decision process, a feature essential to establishing causal links. The aim of this study is to find the causes of restricted group phenomena for majority voting in Checkers, considering agent and group variables in isolation, and try to relate them to similar behavior in other tasks. The term "homogeneous" here refers to the use of copies of the same Checkers game playing program, whereas a "heterogeneous" group would use different programs.

The remainder of this work is organized as follows: Section 2 presents previous work on this and associated problems, and how they relate to this study. Section 3 explains the method used in conducting the investigation. Section 4 presents the experimental results and relevant discussion. Finally, Section 5 provides a summary of the findings and some concluding remarks.

2 Related work

Recent work by Thiele and Althfer [11] sought to build theoretical justification for the independence assumption in majority voting systems, analyzing a situation where the group of experts had even size and was split in independent couples that had dependencies within themselves, on a simple subtraction game. The study was grounded on a modern theoretical framework from Social Sciences, and found a positive relation between group size and performance, but most specifically, that a negative correlation within couples had a positive effect on performance, linking variability of group decisions to performance. However, a causal relation between the observed variability and group performance was not established.

The works of Spoerer et. al. [9, 10] on majority voting for homogeneous and heterogenous groups in Chess provide a starting point for this work, as they confirm the positive effect of group size in a more complex game. The point of attention in this case is that such positive effect was not monotonic and grew in an apparent decreasing rate, suggesting that the group behavior was being steered towards an expected value given by the collective decision process. The construction and implications of such decision process is the object of study in this work.

Obata et. al. [6] analyzed majority voting in the game of Shogi, applying a normal-distributed random variation to the move evaluation function of Shogi program *BONANZA*. Both homogeneous and heterogeneous setups were tested, with considerable advantage being achieved by group consultation when compared to using a single agent. An explanation is hinted for such advantage, given in terms of the relative ability of each member of the group, which is expressed as a probability of selecting a "correct answer", i.e., an advantageous move. With the probability of an advantageous move by the group majority being a function of the members' probabilities, if it exceeds 0.5, then the group consultation becomes advantageous. Hoki et. al. [3] applied the techniques employed in [6] with further improvements to develop *Akara 2010*, a distributed Shogi playing system that could achieve the unprecedented feat of defeating a professional human player in a public game. Later, Hoki et. al. [4] conducted a similar investigation for Chess problems, but exchanging majority voting for *optimistic selection*, with a noticeable increase in performance when compared to single agent Chess problem solving.

Alternatively, Althfer [1] analyzed *multiple-choice* group consultation systems, where the decision process is done at two steps: first a group of computer programs make a move proposal each and finally a human selects one among them, making the final decision. Motivated by the success of their proposed *3-hirn* systems (a particular case of multiple-choice) in previous works with Chess, the authors applied 3-hirn to the game of *Clobbers*, also with positive results.

An understanding of the conditions necessary for group superiority was explored much earlier by Lorge and Solomon [5], where the existence of possible emergent elements in problem solving by groups was examined, to a negative conclusion. This means that a "correct answer" for a given problem cannot be expected to be produced by a group in which no member is able to individually produce such answer. Furthermore, they conclude that the probability of an individual member of the group producing the "correct answer" is dependent not only on its skill, but on the problem itself.

Additionally, the work of Sato et. al. [7] proposed a mathematical proof for expectation of performance on a majority consultation algorithm with random noise. A dependency relationship was indicated between the performance probability distributions of the original single evaluation function vs. the consultation function. Expected improvement of the consultation function should offset any reduction caused by the use of weaker agents (the ones working under noise) for the majority voting to be advantageous. The study concluded that effectiveness of the majority consultation algorithm depends on the game, but without a causal link between the consultation function and the expected performance.

3 Method

A procedure was developed for testing homogeneous group performance in Checkers, with the goal of observing the connection between the player's decision process, the group behavior and the expected performance. A set of important factors are taken into consideration, which are described in the following sections.

3.1 Non-determinism

Since the agents are game playing programs, a fundamental aspect to note is the matter of *non-determinism*: if all agents in a homogeneous group behave in a deterministic way, all their playing decisions will be the same and there will be no difference to a single player. Game playing programs (henceforth "game engines") typically include a form of non-deterministic play into its decision process, using a *Pseudo-Random Number Generator* (henceforth RNG). Therefore, the RNG choice is an important player property. Each RNG is associated with a specific probability distribution.

3.2 Initiative

Another important aspect is the *initiative.* The player that does the initial move may have an advantage or disadvantage, depending on the game. This fact adds a scoring bias into the game and its knowledge creates asymmetric behavior between the players, since one has to compensate the scoring bias to win. Upon examining statistics from professional Checkers tournaments[1], it was noted that this game has a bias against the starting player, like *Reversi (Othello)*, despite being a draw on perfect play. The unbalance is such that the red (starting) player has an approximate winning record of 17% of the matches, 50% of draws and the remaining 33% of loses. Such factor must be taken into account when analyzing performance expectation. The game bias also serves as a pivot, from which player decisions can be separated into beneficial or detrimental.

3.3 Prior variables

With this in mind, the following group and agent properties were observed:

- *Group size*: the number of player agents in a given side.
- *Player skill level*: comprises the search depth measured in plies, and the time available to make a move. Since both were tied together, only the number of plies will be mentioned.
- *Player color*: the side the player takes in the game. The color determines the one who makes the initial move.
- *Player RNG distribution*: determines the source of non-determinism for the player, described by a distribution probability.

---

[1] http://www.usacheckers.com/internationalmatchresults.php

Each property is taken in isolation, changing its value while locking the other properties into specified values. Relationships can then be uncovered by analyzing game results in combination. The next section describes in more detail the experimental setup and methodology that allows drawing some conclusions.

3.4 Experimental setup

The game engine of chosen for this study was *Samuel* [2], an open-source Checkers program based on the GUI Checkers [3] game engine. Samuel was chosen because it was the only open-source program on the higher-than-average quality group of Checkers systems. The open-source characteristic is necessary for understanding the player's decision process and to be able to change the RNG behavior.

Homogeneous group play was implemented by *forking* (duplicating) running instances of *Samuel* for independent move selection, passing as input the current game board configuration. The move to be played in the group's turn is decided by simple majority voting. When no majority exists, the first vote is selected as decisive. The independence condition was checked by running part of the experiments using separate Samuel processes and comparing the results for discrepancies. This game engine uses the RNG produced numbers to select a starting branch for minimax searches.

Experiments were conducted by playing a fixed number of 10 matches and computing the *winning rate* for each player color, by the following formula:

$$winning\_rate = \frac{(num.wins + 0.5 * num.draws)}{num.matches} \tag{1}$$

Each experiment was run with a different setting of group size, player skill level and RNG distribution. They were repeated 100 times each, and the collected statistics were analyzed. The prior variables were configured as shown in Table 1.

Table 1. Experiment settings

Variable	**Range/Types**
Group size	1, 3, 4, 5, 8, 10, 15, 20 players
Player skill level	1 to 12 plies
Player color	$RED, WHITE$
Player RNG distribution	$Uniform, Gaussian$

The Uniform distribution was set for the interval of [1, n_moves], where n_moves is the possible number of moves for the moving player in each turn. The Gaussian distribution was set with a mean of $\frac{n\_moves}{2}$ and a standard deviation of $\sqrt[2]{\frac{n\_moves}{2}}$. Values below 1 or exceeding n_moves were capped to those two limits respectively. The Gaussian distribution parameters were selected to make them centered regarding the move set.

Observations were primarily done taking the red (starting) player as reference, since increases in performance can be more promptly checked for being

---

[2] https://github.com/johncheetham/samuel

[3] http://www.3dkingdoms.com/checkers.htm

due to beneficial bias induced by a variable change, rather than reflecting a reinforcement of the game bias. The professional U.S. Checkers tournament statistics were taken as reference for the bias, with a winning rate of 0.425 for red and 0.575 for white. The winning rate mean for each set of 10 experiments is used as performance expectation measure. Variations in the winning rate indicate a beneficial or detrimental change in playing ability for the agent or group. Despite being a high granularity performance measure when compared to move scoring, the winning rate was chosen due to being the only one that can be measured reliably in this setup, since moves are scored by a heuristic evaluation function in the *samuel* program. Furthermore, it would be necessary to evaluate the entire game tree (or simulate it through a program) to obtain an absolute quality measure for a move, which is not feasible for the game of checkers.

4 Experimental results

Firstly, the statistics for a single player game in both sides were collected and the results are shown in Figure 1.

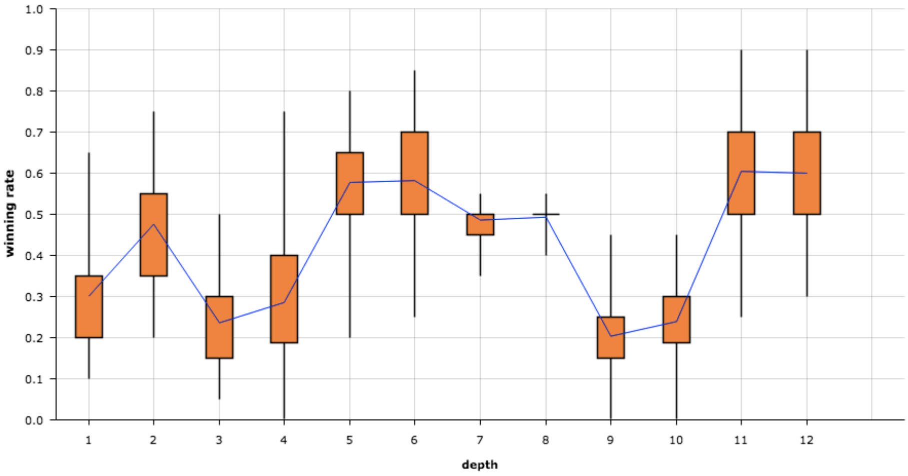

Fig. 1. Mean winning rate for the red (starting) player using the uniform RNG, as function of the skill level, measured in search depth level (plies). The boxes represent the 25-75 percentile range.

The results indicate that there is a changing tide of performance, which is dependent on the skill level of both players. However, there is no indication of correlation between skill level and performance, when both players have the same skill level.

Next, Figure 2 shows a comparison of performance for groups of up to 20 players on the red side, playing against a single one on the white side.

This time, there is a clear convergence of the group players towards a better performance expectation between 4 and 6 plies of search depth, but for a worse

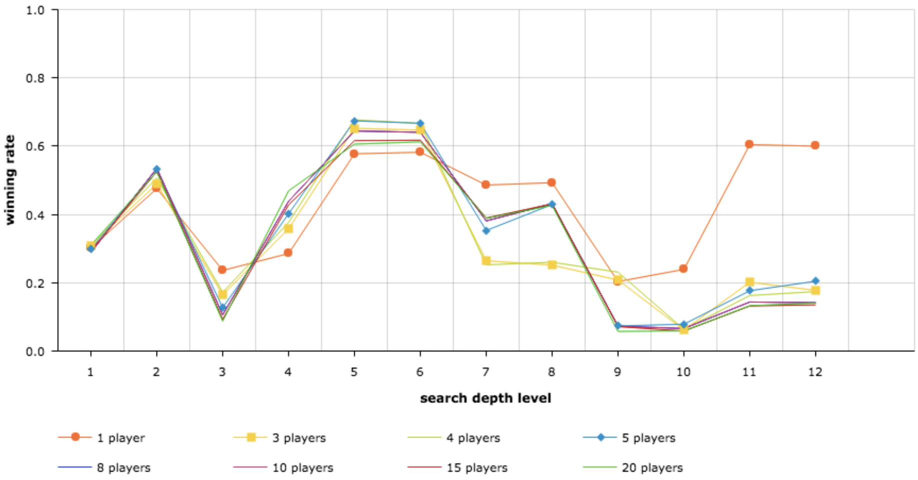

Fig. 2. Mean winning rate for groups with 1, 3, 4, 5, 8, 10, 15 and 20 players on the red side using the Uniform RNG, as function of the skill level, measured in search depth level (plies).

performance expectation from 7 to 12 plies. This indicates a higher probability of selecting a beneficial or detrimental move for the red player in the respective skill levels. Using groups of up to four players results in unstable play, which performs similarly to larger groups except between 7 to 9 plies, where it behaves either as single player or takes a sharp drop in performance. The reasons for this phenomena are still under study, but the move selection data indicates a reduced number of viable moves for the red player under the 7 to 9 plies condition. This makes a 3-4 group majority with first selection more likely to select a bad move at random if there are only two options available, whereas a single player or a larger group would select either with equal probability. Changing the program to select a move at random when there is a tie in the voting is likely to change this.

The white player collects opposite benefits and losses when playing in groups, as shown in Figure 3. However, the gains and losses are not symmetrically opposite, due to the game bias inducing a different level of advantage for each side, which is not changed even when using groups.

The change in performance expectation is consistent for both players and indicate a bias in its distribution that is dependent on the skill level of the players. However, such bias might be tied to the players decision process itself. This hypothesis can be tested by changing the RNG distribution of one of the players.

In Figure 4 the performance expectation of a player using a *gaussian* RNG is compared to one using the *uniform* one, showing an overall performance gain. This indicates the possibility of steering the player in a beneficial way, specially when informed about the opponent skill level.

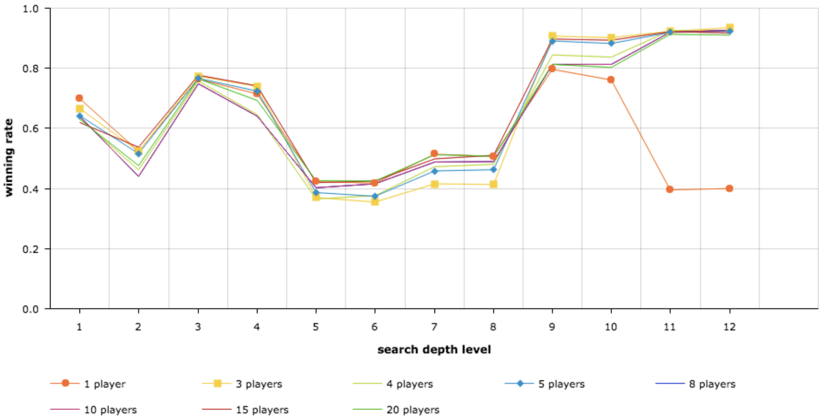

Fig. 3. Mean winning rate for groups with 1, 3, 4, 5, 8, 10, 15 and 20 players on the white side, as function of the skill level, measured in search depth level (plies).

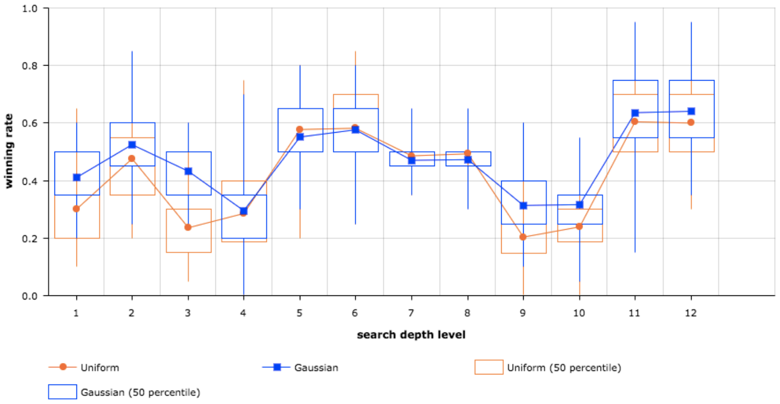

Fig. 4. Comparison of mean winning rates for the red player as function of the skill level, for the uniform and gaussian RNG distributions, measured in search depth level (plies). The boxes represent the 25-75 percentile range.

However, as shown in Figure 5, playing as a group makes the performance expectation converge in a similar way as seen in Figure 2, a phenomena that can be explained as a bias in the game itself, which also provides an explanation for the higher white winning rate in professional human Checkers tournaments.

The experimental results not only confirm the theoretical findings of Lorge and Solomon [5] and Sato et. al. [7], but also point to the game structure as the underlying cause of group phenomena. This can be concluded by observing the convergence of the group play performance w.r.t. the group size, towards a line that is independent of distribution biases in the decision process of each agent (Figures 2 and 5). Thus, the group behavior, although directed by the agent's decision process, is defined primarily by the game structure, which drives the expected performance for a fixed skill level.

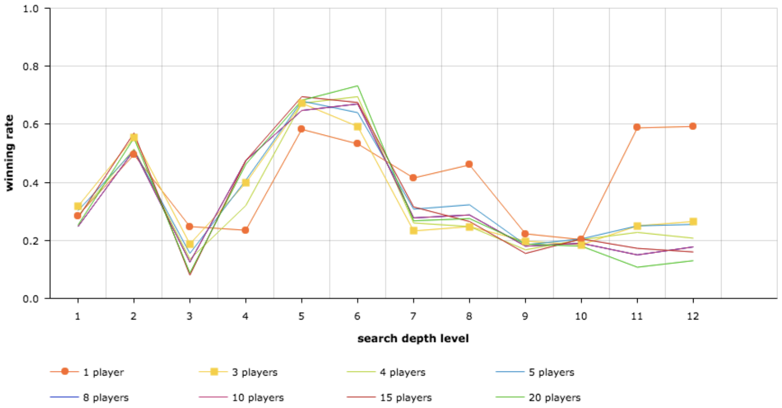

Fig. 5. Mean winning rate for groups with 1, 3, 4, 5, 8, 10, 15 and 20 players on the red side, using the gaussian RNG distribution, as function of the skill level, measured in search depth level (plies).

A simple explanation for the observations could be expressed in the following statement: For a given game and a certain skill level, at any point in the game, there will be a number of "good" and "bad" moves, for which balance is defined by the game structure. When selecting a move through majority voting, the player increases the chance of drawing the dominant type, be it good or bad.

The results also follow the findings of Thiele and Althfer [11] when considering the subtraction game, since for that game there can be only a single good and bad move. However, their extrapolation to games with more than two possible moves is unlikely to occur, given the evidence here analyzed. Interestingly, *3-hirn* consultation reduces multiple move choice to only two options, in which one of them will be better or at least equal to the other. This limits the effect of the game bias and put more weight on the agents' skill.

Furthermore, by changing the underlying distribution used in the decision process, it is possible to steer the performance expectation, specially when information about the opponent skill level is available. As it can be seen in Figure 4, it is possible to obtain a stochastic dominant player, even in a situation of initiative disadvantage.

5 Conclusion

Understanding the underlying reason for group performance phenomena in particular conditions is not often explored and may help to improve performance in known setups by changing agent behavior, or find new ways of combining agents to take better advantage of their characteristics.

In this work, we analyze homogeneous group performance in Checkers, using independent computer playing agents, and observe the connection between the player's decision process, the group behavior and the expected performance. Experiment results point to a causal link between the game structure and the performance expectation, as a function of the player's skill level. Additionally, we found the possibility to steer performance expectation into favorable levels by manipulating the game engine's Pseudo-Random Number Generator (its source of non-determinism), which is particularly effective when the opponent's skill level is known.

Future works include testing other games under the same experimental framework, and trying to pinpoint specific properties of the Checkers game that contribute to the bias in the game structure.

References

1. Althöfer, I.: Improved game play by multiple computer hints. Theoretical Computer Science 313(3), 315–324 (2004)
2. Hackman, J.R., Morris, C.G.: Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. Advances in experimental social psychology 8, 45–99 (1975)
3. Hoki, K., Kaneko, T., Yokoyama, D., Obata, T., Yamashita, H., Tsuruoka, Y., Ito, T.: A system-design outline of the distributed-shogi-system akara 2010. In: Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2013 14th ACIS International Conference on. pp. 466–471. IEEE (2013)
4. Hoki, K., Omori, S., Ito, T.: Analysis of performance of consultation methods in computer chess. Journal of Information Science and Engineering 30(3), 701–712 (2014)
5. Lorge, I., Solomon, H.: Two models of group behavior in the solution of eureka-type problems. Psychometrika 20(2), 139–148 (1955)
6. Obata, T., Sugiyama, T., Hoki, K., Ito, T.: Consultation algorithm for computer shogi: Move decisions by majority. In: International Conference on Computers and Games. pp. 156–165. Springer (2010)
7. Sato, Y., Cincotti, A., Iida, H.: An analysis of voting algorithm in games. In: Computer Games Workshop at European Conference on Artificial Intelligence (ECAI). pp. 102–113 (2012)
8. Schaeffer, J., Burch, N., Björnsson, Y., Kishimoto, A., Müller, M., Lake, R., Lu, P., Sutphen, S.: Checkers is solved. science 317(5844), 1518–1522 (2007)
9. Spoerer, K., Sirivichayakul, T., Iida, H.: Homogeneous group performance in chess. Procedia Technology 11, 1272–1276 (2013)
10. Spoerer, K.T., Okaneya, T., Ikeda, K., Iida, H.: Further investigations of 3-member simple majority voting for chess. In: International Conference on Computers and Games. pp. 199–207. Springer (2013)
11. Thiele, R., Althöfer, I.: An analysis of majority systems with dependent agents in a simple subtraction game. In: International Conference on Computers and Games. pp. 202–211. Springer (2016)
12. Triplett, N.: The dynamogenic factors in pacemaking and competition. The American journal of psychology 9(4), 507–533 (1898)
13. Webber, R.A.: The relation of group performance to the age of members in homogeneous groups. Academy of Management Journal 17(3), 570–574 (1974)
14. Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W.: Evidence for a collective intelligence factor in the performance of human groups. science 330(6004), 686–688 (2010)