Mistake patterns among human Go players: Insights from AI

Carlos G Urzúa-Traslaviña[1], Quentin Rendu[2]
[1]*University of Groningen (RUG)*
[2]*Hamburg University of Technology (TUHH)*

Abstract

Artificial Intelligence (AI) is now routinely used by Go players to review their games. Analyzing individual mistakes helps players identify their weaknesses. However, deriving generalizable insights requires a broader analysis of mistake patterns.

In this study, 100,682 AI-scored amateur and professional Go games are studied to investigate mistake patterns. Three different ranks are examined, ranging from low-level amateurs to top professionals. Various move features such as height, distance to previous move, and adjacent stones are analyzed to gain a deeper understanding of these mistake patterns.

The key findings are as follows: (1) a noticeable improvement in opening performance among professional players since 2017; (2) a significant performance gap in the endgame between professionals (who exhibit near-optimal play) and amateurs; and (3) areas for improvement in tactical skills among amateurs, particularly in first-line and sacrificial moves.

Keywords: Go, Baduk, Weiqi, AI, Katago, Mistake distribution

I. Introduction

Go, with its simple rules and emergent complexity, has attracted players for hundreds of years. With the advent of personal computing, expert systems for perfect information games such as Chess, Backgammon and Go became a reality. Initially, these systems relied mostly on pre-existing datasets of human games and on human designed evaluation functions. As such it was not expected they could develop novel methods of play. This changed with the development of AlphaZero (Silver et al., 2016; Silver et al., 2017) an expert system that achieved superhuman performance in perfect information games purely from self-play. Nowadays, derivative AI expert systems (e.g. KataGo) are used by amateur and professional go players to study their games. The analysis of mistakes (i.e. moves associated with a significant point loss) is of particular interest to improve one's skill (Shin et al. 2021). Certain types of mistakes, such as shallow tactical errors in Chess, are correlated to a player's strength (Saariluoma, 1992). Mistake distributions give hindsights on a player's performance during one game. Such information has been used in Chess to determine the rank of a player (Ferreira, 2012), or to evaluate the inflation of the Elo scale (Regan et al., 2011). In Go, it has recently been used to design anti-cheating detection algorithms (Egri-Nagy et al., 2020; Barollet et al., 2023).

An important challenge when using AI to review a game is to understand the reasoning behind a mistake, especially for beginners. Mistakes are often prioritized by their estimated point loss. However this point loss reflects the expected outcome of the game if two AI players continued playing and not the loss when playing a human opponent of the same rank. The previous observation that AI win rate does not reflect the actual win rate of a beginner supports this (Rendu, 2023). We wondered whether AI scoring of human matches could be used to discover recurrent human mistakes that are most advisable to improve upon by amateur human level play.

In this article, we study 100,682 AI scored games from players of ranks 12k (beginner), 1d (amateur) and 9p (top rank professional) obtained from the previously compiled Analyzed Kifu dataset. We then annotate moves based on their turn number, line height, shape with adjacent stones, and distance to previous move. We show that these features capture previously known facts about human matches. Finally we use this annotated dataset to identify recurrent mistakes of these three ranks of players.

II. Methods

1. Data curation

Scored human games were extracted from the Analyzed Kifu Database (Rendu, 2023). Specifically, we selected even matches analyzed with Katago v1.12.4 (Wu, 2019), using the neural net ”b18c384nbtuec-20221121b” with 5 visits. We study three ranks: 12k matches with 1200s main time (Fox server), 1d matches with 1200s main time (Fox server) and 9p matches with 300s main time (go4go).

After parsing the moves and metadata of these games from the SGF files we applied the following quality control. Since we are interested in human moves we removed games played by a bot (“GoogleDeepMindAlphaGo”). We then observed that resignation of beginner matches often occurs long after the moment when the game is realistic to turn. While moves during this end phase may not represent attentive play, accurate evaluation of a game is a skill that beginners may lack. As a compromise we identified moves where the expected AI winrate was less than 5% and the estimated point loss by AI was bigger than 100 points. We believe this threshold conservatively discards moves where the inevitable outcome of the game is clear even for a beginner. We ignored all moves that occurred later than the earliest instance of such desperate moves in all games. This resulted in a remaining total of 20,094,055 moves from 100,682 games from 41,282 players. Since the frequency of games after move 350 is minimal and a peak of games was observed in 1d matches at that turn, only the first 350 moves of each game are analysed in this study (Figure 1).

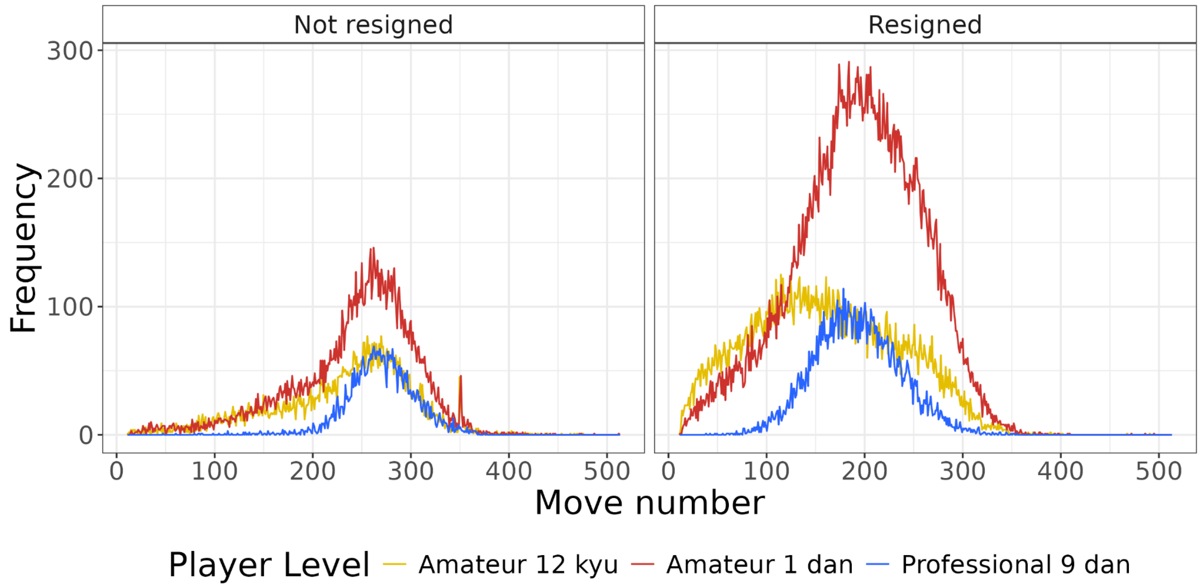

Figure 1: Frequency of moves (y axis) by turn number (x axis) in the three studied ranks for resigned and not-resigned games. A peak observed in 1d games at move 350. We believe this reflects a rule in Fox server where counting can be forced on move 350.

Point loss is defined as the score lead difference between the current and the last move. Playing the best move according to AI should lead to a point loss of near 0. All the other moves should ideally have a negative value. Due to the low number of visits used in the scoring of the selected games, some played moves are better than AI choice, leading to a small subset of positive point loss moves. We decided to discard moves with positive values for the rest of the analysis.

2. Definitions

Instead of a global classification of moves that are mistakes, when studying a group of moves we indicate the threshold for mistake moves as the 25% percentile of that group. For example, when studying the group of moves of 1d games at turn 150 the worst ¼ of moves are considered mistakes. But when studying the subset of moves of 1d games at turn 150 that are cuts (see below) the percentile is recalculated for this subset definition.

We annotated the moves with the following features.

- Move height: Distance in empty spaces from the coordinate of the move to the nearest border of the board..
- Tenuki: Boolean that states if the current move is more than 5 empty spaces away from the previous move in either the x or y axis.
- Move types: A played stone is in connection with 4 orthogonal adjacent positions. We define two types by contact with stones of the same color regardless of the contact with enemy stones: Extension (1 friendly stone), Connection (>1 friendly stones). We define five types when there is no contact with friendly stones by counting liberties: Throw-in (0 liberties), Ko/Sacrifice (1 liberty), Cut (2 liberties), Attachment (3 liberties), Placement (4 liberties). These rules are adapted for moves at the border

and corners of the board. Border: 3 liberties at border is Placement and 2 liberties is Attachment, cut does not exist. Corner: 2 liberties at corner is Placement. cut and attachment do not exist.

We verified that these features behave as expected during the course of the game. For example, moves in the 3rd and 4th line are popular in the first moves (opening) while moves in the 1st line are more frequent at the endgame (Fig 2A). In the literature, the move index at which the endgame starts has been reported to be 162 on average, with a standard deviation of 19 (Li et al., 2019). This is consistent with our results, which show a large increase in 1st line move frequency between move index 100 and 200 (Fig 2A). Similarly Tenuki is popular during the opening and their frequency steadily increases as the game advances (Fig2B). For the case of move types, cuts and extensions are frequent throughout the game, placements and attachments are more frequent early and ko/sacrifices are more frequent late in the game.

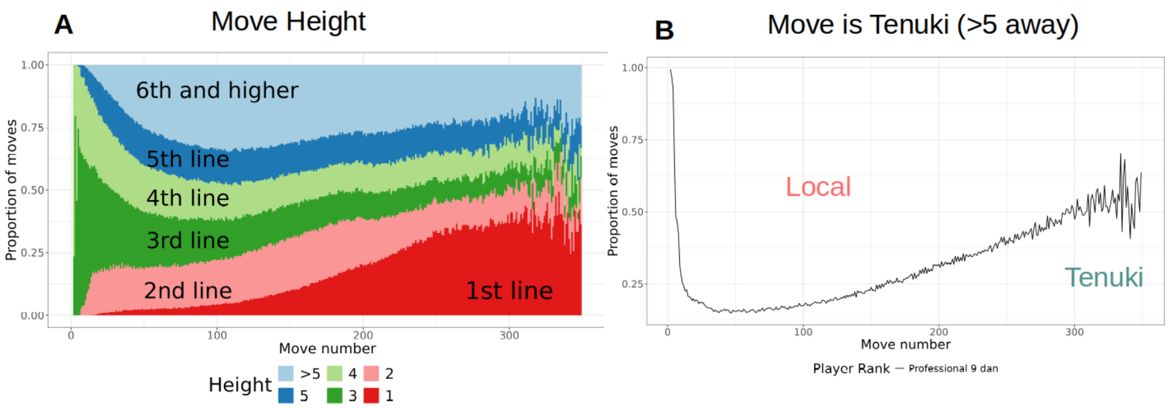

Figure 2: A) Proportion of moves (y axis) by turn number (x axis) in professional games colored by move height. B) Proportion of tenuki moves (y axis) by turn number (x axis) in professional games.

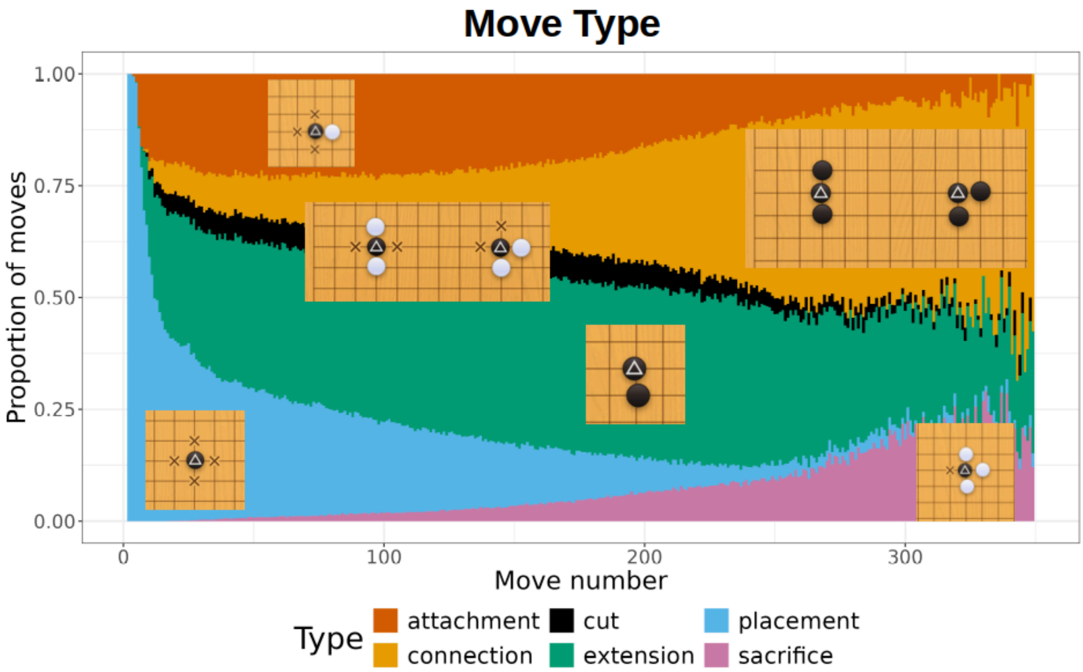

Figure 3: Proportion of moves (y axis) by turn number (x axis) in professional games colored by move type. A representative example of the type of move is showcased.

In the results section, the mistake distributions are presented for different player ranks and at different epochs (pre and after AlphaZero). Patterns are analysed to derive insights and knowledge about the common mistakes played by amateurs with respect to professional players.

III. Results

1. Mistake patterns before AlphaZero (before 2017)

Since amateur games may not reflect traditional play we first explored distributions derived from the professional game dataset. First, only games played before the development of AlphaGo were selected. We observed that the threshold for mistake is lowest (worse) between moves 100-150 and highest after move 250. This shows that for pros the endgame is closest to optimal play, followed by the opening, with the middle game being the most difficult according to Katago. Additionally we observed that moves closer to the edge of the board are rated worse (Fig 4A). Since we do not have the scores for alternative moves at the studied positions two explanations are possible: moves closer to the edge are easier to get wrong and/or when these moves are played there is often a better move elsewhere.

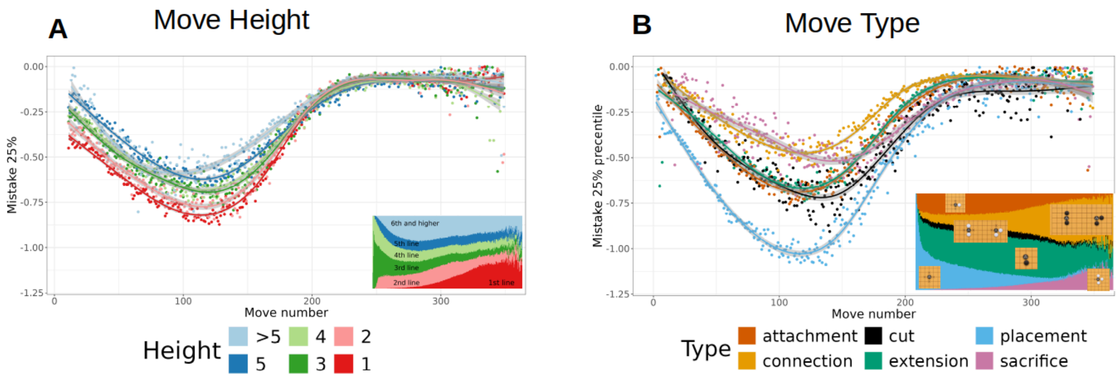

Figure 4: A) Mistake threshold (25% percentile of point loss) (y axis) for all moves of a specific turn (x axis) and line number are plotted. Inset shows proportions of moves. B) Mistake threshold (y axis) for all moves of a specific turn (x axis) segregated by move type are plotted. Inset shows proportions of moves.

When exploring the move type we again observe a similar inverted bell shape, but the variability between move types is higher than for move height (Fig 4B). Attachment, cut and extension all have comparable mistake thresholds. Connection shows the smallest mistake threshold (sacrifice mistake threshold is lower before move 100 but with a very low occurrence), whereas placement is associated with the highest mistake threshold. A mistake in placement can be interpreted in two ways: either the best move was a placement but at another intersection, or the best move was of a different type.

To narrow down the influence of tenuki on the mistake threshold, two distributions are plotted in Figure 5 for each move type: one for moves played locally and one for tenuki moves (distance > 5 to previous move). Overall, one can see that tenuki is associated with higher mistake threshold than local moves except for the placement and sacrifice move types. This suggests that pros are more effective at responding to local positions than for selecting a tenuki move. Again we cannot determine in these cases if a local move was better, or if the best move was a different tenuki move.

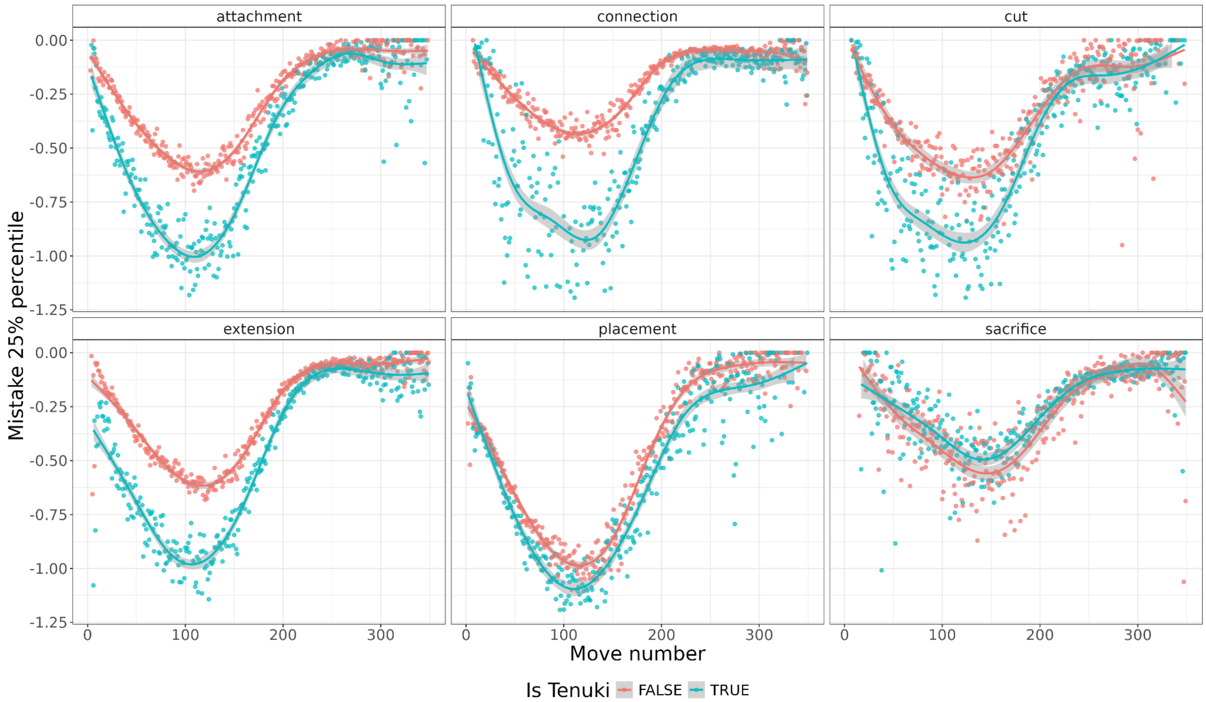

Figure 5: Mistake threshold (25% percentile of point loss) (y axis) for all moves of a specific turn (x axis) segregated by tenuki (blue) not tenuki (red). Inset shows proportions of moves.

2. Influence of AlphaZero on professional players mistake patterns (after 2017)

We then proceeded to compare Pro matches before and after the development of AlphaZero. We immediately observed a noticeable improvement in the opening of the game after the development of AlphaZero (Fig 6). Additionally we observed a small shift in the inverted bell curve suggesting that the hardest turns of the game occur later. A possible explanation is that changes in the opening affected the difficulty of the middle game.

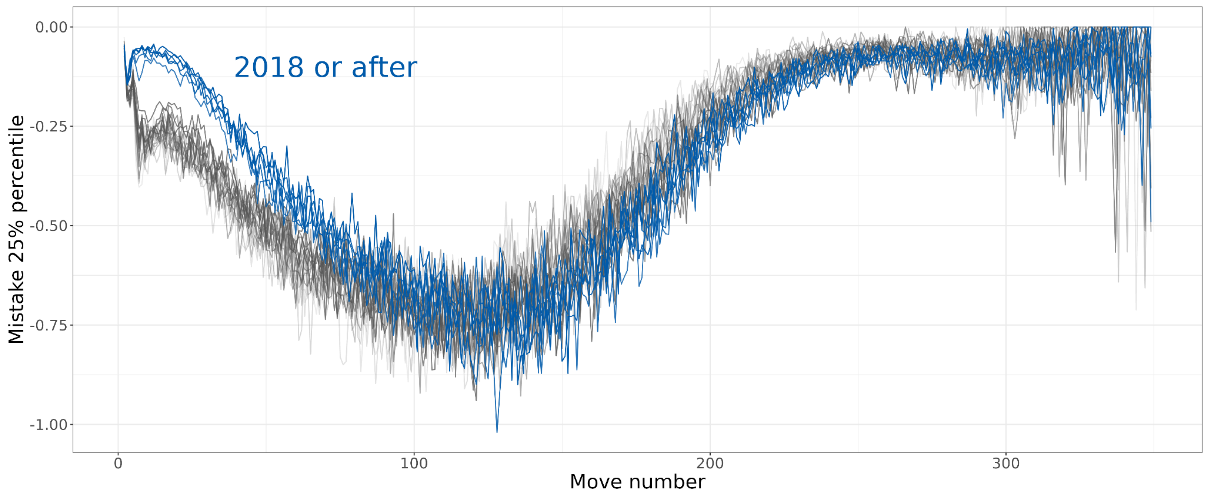

Figure 6: Mistake threshold (25% percentile of point loss) (y axis) for all moves of a specific turn (x axis) segregated by year of match.

To further investigate the improvement of professional players performance during the opening, the difference in mistake threshold is plotted for every turn in Figure 7. Three different move types, common in the opening, are depicted: attachment, extension and placement. The variability is quite high, especially for placement moves. No clear trend is visible above move index 100. A clear positive region can be seen for all plots for move index 50 and below. This indicates that the improvement in the opening is distributed along all types of moves rather than specific to one kind of move.

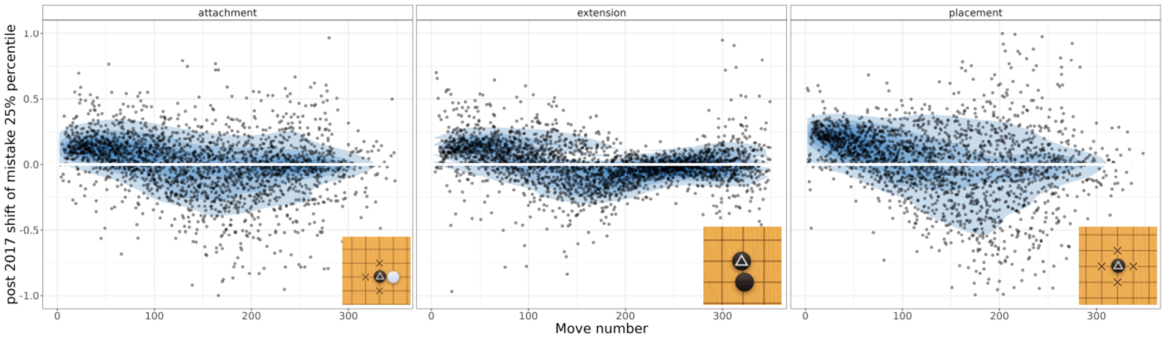

Figure 7: Delta of mistake threshold before and after 2017 (y axis) for all moves of a specific turn (x axis) segregated by most common move type in opening.

3. Influence of player rank on mistakes patterns

After establishing a baseline with high quality pro games we proceed to compare the distributions of mistakes in lower ranks. Both live games (for professional players) and online games (played on Fox Go server, for 12k and 1d amateur players) are used in this section.

First, the global mistake threshold is plotted along the move index in Figure 9A. The characteristic inverted bell shape is seen for the three datasets, although the amplitude depends on the player rank. As expected, low rank (12k) is associated with the lowest mistake threshold, which reaches -6.2pt around move 150. Players with a slightly higher rank (1d) tend to make smaller mistakes, with a minimum of -4.1pt around move 150. Professional players' mistake distribution reaches a minimum of -0.75pt around move 125. This shows that the highest difficulty of the game occurs earlier for pro matches. Another noticeable difference is the high amount of 1 point mistakes in the endgame for 12k and 1d players (Fid 8A-B). This suggests that endgame proficiency is not common amongst low and mid level amateur players.

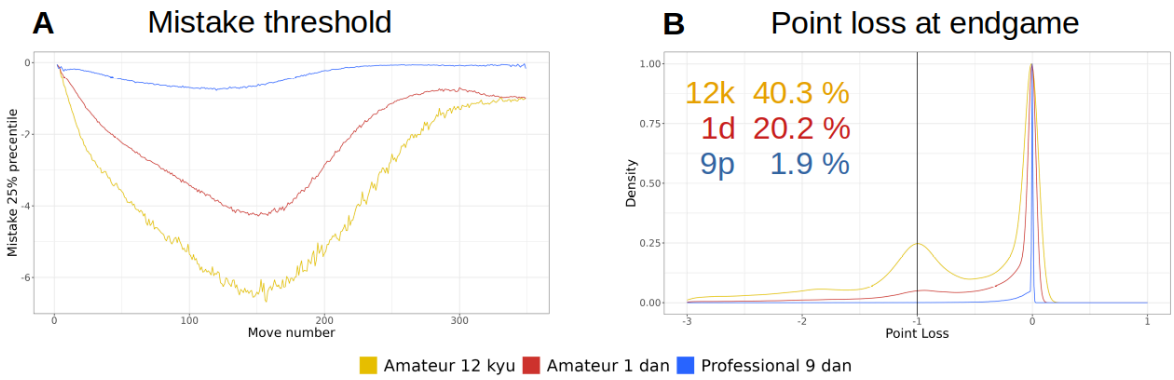

Figure 8. A) Mistake threshold (25% percentile of point loss) (y axis) for all moves of a specific turn (x axis) segregated by player rank. B) Normalized density plot of point loss of endgame moves (turn 250-350). Inset shows the proportion of moves per rank that have a point loss less than -1.

We then compare the frequencies of different move types and height throughout the match Fig 9. The frequencies are very similar irrespectively of the players' rank. Two notable exceptions can be spotted. The first one concerns the 1st line moves and high (>5 line) moves played in the opening phase (move index < 100) and middle game phase (100-200) respectively. One can see that the amateur players frequencies are higher than professional players frequencies. First line moves are probably premature in these amateur games. A similar explanation goes for premature high moves observed in the middle game. The other exception lies between the opening and the middle game, around move index 100, and concerns second line moves. These are less frequent amongst amateur players who tend to play them later in the game. Two line moves are generally of great importance for the endgame, they should be played as early as possible but not too early. The distribution observed here could also be an indicator of the better timing abilities of higher ranked players.

On the bottom part of Figure 9 are plotted the frequencies for the different types of moves. Once again they are very similar irrespectively of players' rank. Some deviations can be observed, 12k players use less sacrifices and placements, and professional players use less extension moves and favor late connections. Note that with the definition given in the Methods section, the sacrifice moves include taking (or retaking) a ko. We also observed that pros tenuki more than amateurs (Fig 10).

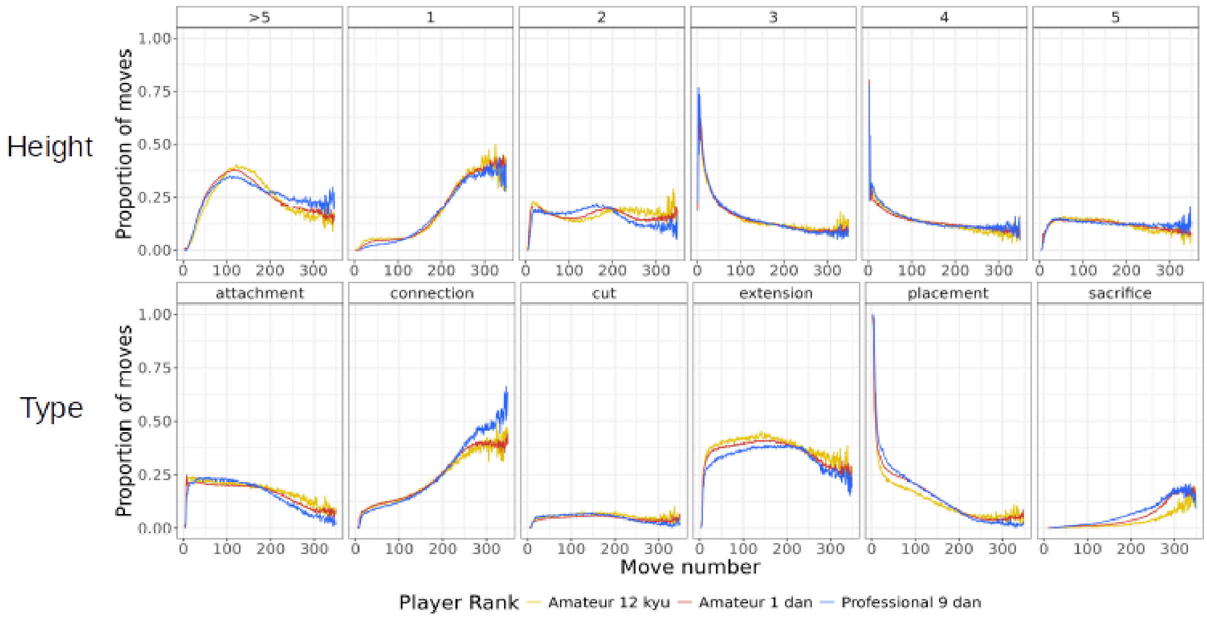

Figure 9: Top) Proportion of moves (y axis) colored by player rank and segregated by move height. B) Proportion of moves (y axis) colored by player rank and segregated by move type.

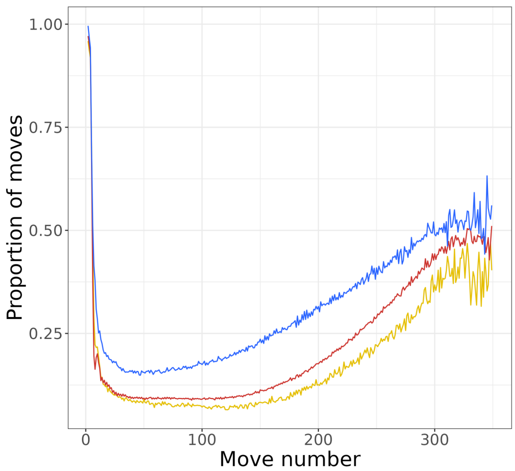

Figure 10: Proportion of tenuki moves (y axis) colored by player rank. Pros tenuki more than amateur players.

We then compare the mistake thresholds of move types at different moments of the game between 2 ranks by subtracting the thresholds calculated for professional players from the mistake threshold calculated for amateur players. This revealed that both 1d and 12k players are not uniformly worse than pros at all moves. Moves at the first line were rated worse than moves at other lines suggesting that amateur players are especially weak at 1st line moves irrespective of move type.

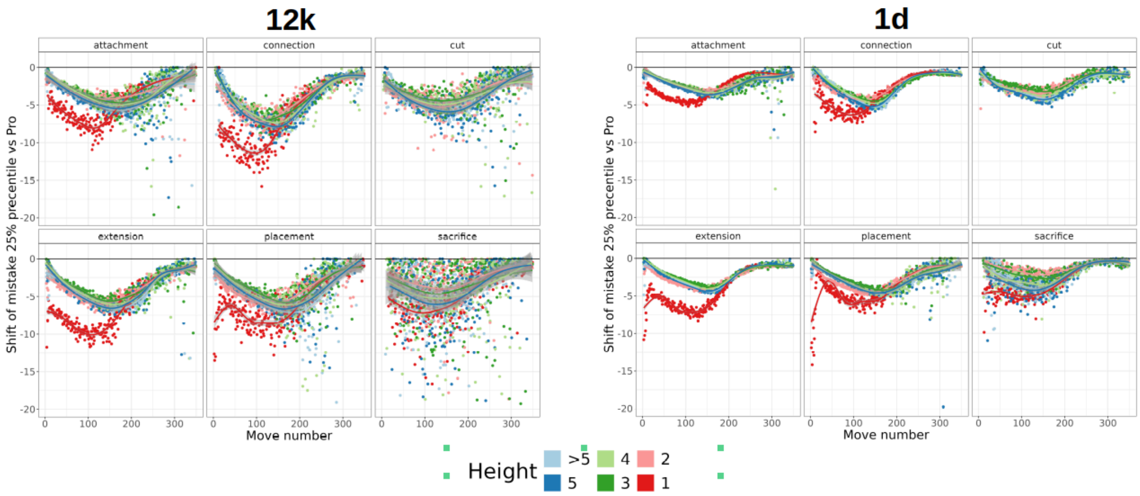

Figure 12: Delta mistake threshold of amateur players (1d/12k) relative to professional players. Colored by line height and segregated by move type. Note that cut cannot occur at 1st line according to our definition.

IV. Discussion

In this article, we studied 100,682 AI scored games from professional and amateur players of ranks 12k (beginner), 1d (amateur) obtained from the previously compiled Analyzed Kifu dataset. After annotating features to segregate move types by local shape and by using professional player matches as baseline we confirmed that the opening changed after AlphaZero towards moves with better AI score. We were able to identify the following recurrent mistake patterns for amateur players. 1) Higher proportion of premature 1st line moves irrespective of move type 2) higher proportion of 1 point moves at endgame 2) Lower frequency of tenuki suggesting bias towards local play 3) Avoidance of sacrifice and placement moves (12k) moves and overreliance on extension moves (12k and 1d).

This analysis highlights the transformative role of AI in Go. Through individualized use of AI, professional players have adapted to new strategies, resulting in fewer errors, particularly in the opening phase. Compared to professional players, amateurs show areas of opportunity in tactical situations like first-line moves as well as in timing abilities, such as second line moves during the first stage of the game. For the case of professional players only the endgame appears to be near optimal play.

V. Conclusions

We anticipate that this analysis can be improved by the exploration of AI systems that mimic human mistake distributions such as newly developed Katago SL agents. For example we could then determine rank-reasonable alternatives for moves that we considered a mistake. This would unlock the ability to indicate if there is a recurrent better play for some mistakes. Another possibility of this analysis is the prioritization of mistakes not by point loss but by frequency in equivalently ranked players. For example by focusing on 1st and second line mistakes and 1-point loss moves at endgame. We hope analyses of amateur games can contribute to more effective game review using AI.

References

Barollet, T., & Le Duc, F. C. (2023). A statistical analysis of amateur go players to assist AI-cheating detection in online go communities. *Journal of Go studies*, *17*(2), 89-106. doi: 10.62578/980526

Egri-Nagy, A., & Törmänen, A. (2020, November). Derived metrics for the game of Go–intrinsic network strength assessment and cheat-detection. In *2020 Eighth International Symposium on Computing and Networking (CANDAR)* (pp. 9-18). IEEE. doi: 10.1109/candar51075.2020.00010

Ferreira, D. R. (2012). Determining the strength of chess players based on actual play. *ICGA journal*, *35*(1), 3-19.

Li, X., Lv, Z., Wang, S., Wei, Z., Zhang, X., & Wu, L. (2019). A middle game search algorithm applicable to low-cost personal computer for go. *IEEE Access*, *7*, 121719-121727. doi: 10.1109/access.2019.2937943

Regan, K. W., Macieja, B., & Haworth, G. M. (2011). Understanding distributions of chess performances. In *Advances in computer games* (pp. 230-243). Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-31866-5_20

Rendu, Q., & Hamburg, T. U. (2023). Go players should not trust AI win rate. *Journal of Go studies*, *17*(2), 61-88. doi: 10.62578/476299

Saariluoma, P. (1992). Error in chess: The apperception-restructuring view. Psychological research, 54, 17-26. doi: 10.1007/bf01359219

Shin, M., Kim, J., & Kim, M. (2021). Human learning from artificial intelligence: Evidence from human Go players' decisions after AlphaGo. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43, No. 43).

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484-489. doi: 10.1038/nature16961

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, *550*(7676), 354-359.doi: 10.1038/nature24270

Received: 30, Apr, 2025
Accepted: 15, May, 2025