JOURNAL OF
EMERGING INVESTIGATORS

Article

Effects of material advantage and space advantage on the Komodo and Stockfish chess engines

Rohan Kaushikan[1], Won Park[2]
[1] Menlo-Atherton High School, Atherton, California
[2] Department of Computer Science and Engineering, University of Michigan, Ann Arbor, Michigan

SUMMARY

Chess engines are computer programs built to play chess. Currently, chess engines are significantly better at chess than the highest ranked human chess players. The utilization of chess engines is the most widely accessible, reliable, and efficient way to study chess. Because of their overwhelming chess capabilities, chess engines are able to aid chess players ranging from amateurs to Grandmasters. A key feature of chess engines is their ability to analyze a chess position and provide a numerical evaluation that describes the degree to which a side is winning. However, different chess engines employ different methods of evaluating a position. Some engines highly value "material advantage" and will greatly favor the side that has captured more pieces. Others put great importance in "space advantage" and will generally try to move their pieces from starting position to positions where they control more squares as early as possible. This study examines how two top-rated chess engines, Stockfish and Komodo, differ in their methodologies for evaluating a chess position. We hypothesized that Komodo's evaluation would change more than Stockfish's when material was lost, and that Stockfish's evaluation would favor the side with more available advancing moves than Komodo. We discovered that Komodo indeed valued material advantage significantly more than Stockfish. However, the two engines did not differ significantly in their evaluations of space advantage. These findings have important implications for chess players, as they allow for accurate imitations of an engine's play style, which could serve to more effectively increase performance.

INTRODUCTION

In 1997, IBM computer engine Deep Blue defeated Russian world chess champion Garry Kasparov in a chess match (1). Chess engines, or computers that play chess automatically, were initially unable to compete at the highest level when they were first developed (2). However, they rapidly improved via millions of practice games, machine learning, and the use of neural networks, with each chess engine developer employing different techniques to create a strong engine (3, 4). Once engine superiority over humans, marked by Kasparov's defeat, was established, the purpose of chess engines changed to aiding chess players, both amateur and professional (5). The advancement of chess engines gave players a valuable opportunity to improve. Specifically, players could now play full chess games against engines to gain experience, analyze previous games, and identify their inaccuracies and mistakes (6). In addition, players could learn chess fundamentals by practicing tactics and early-game positional patterns (6).

Today, many powerful chess engines exist. For reference, the average Elo (a numerical representation of a player's skill level in chess) among all chess players is about 800. At the time of writing this paper, the highest Elo ever reached by a human is 2882 by Grandmaster and former world champion Magnus Carlsen in May of 2014 (7). However, this rating is dwarfed by Stockfish, widely acknowledged as the strongest chess engine, which holds an Elo over 3500. Given that it is possible to numerically represent the strength of a chess engine, tournaments have ensued to determine which is the most powerful. Currently, using Elo ratings determined by Computer Chess Ratings List (CCRL), the strongest chess engine in the world is Stockfish (CCRL Rating 3564), immediately followed by Komodo (CCRL Rating 3508) (8). These discrepancies between human and computer Elo have made chess engines a reliable tool for improving in chess, rivaling over-the-board training (9).

After inputting a board position, chess engines display a value that indicates which side is winning and the degree to which the chess position favors them. For example, a rating of 0.0 means that the position is perfectly equal, and if the best moves are played by both sides, the game will result in a draw. A positive evaluation means that white has an advantage, while a negative evaluation means that black has an advantage. The degree to which one side is winning is indicated by the distance of the evaluation from 0.0. Both the Komodo and Stockfish engines use the same scale for evaluation and have no definite limits to their evaluations, but evaluations larger than ±50.0 are rare. When the game ends, the evaluation will either display 1-0, signifying white's win, 0-1, signifying black's win, or ½-½, meaning a draw.

Position evaluations vary from engine to engine. Given a position, different engines may provide different evaluations. Due to the artificial intelligence learning process used to build chess engines, the algorithm each engine uses to calculate position evaluations is not known, and so different engines may be influenced to greater extents by certain factors in a board position (10). Two factors under consideration are material advantage and space advantage, which were chosen because they influence a player's ability to control squares on the board and execute attacks on enemy pieces (11). Material advantage is determined by how many points of material each side has captured (queens are worth 9 points, rooks 5 points, knights and bishops 3 points, and pawns 1 point). space advantage is defined as how many possible choices for

https://doi.org/10.59720/23-131

advancing moves a certain side has in a position.

This study examined the Stockfish and Komodo chess engines, and how they are differently influenced by material advantage and space advantage. We hypothesized that Komodo and Stockfish would place different levels of importance on material and space advantage. To test this, we manipulated pieces on the board against each engine, recorded the possible moves for each side, and monitored changes to the evaluation score from Komodo and Stockfish chess engines on Chess.com. We found a significant difference between the amount of importance Stockfish and Komodo place on material advantage. However, piece forward mobility, or space advantage, did not have a significant impact on either Stockfish or Komodo's evaluation. These findings suggest that Komodo values the objective of capturing enemy material while retaining one's own pieces more than Stockfish does, and that neither engine places significant importance on space advantage.

RESULTS

In order to test the hypothesis that Stockfish would value material advantage and space advantage differently than Komodo, we set up and conducted experiments on Chess.com's built-in analysis feature (12), which has the capability to use both Stockfish and Komodo as its back engine. For each back engine used, chess positions were generated through the Explorer feature on Chess.com, and evaluations were recorded before and after pieces were removed.

Material Advantage

To determine whether Stockfish and Komodo's evaluations were impacted differently by changes to material advantage, we analyzed evaluations at 11 positions as pieces were removed in a total of 80 tests. Loss of material significantly affected Komodo's evaluation of a position (6.720 ± 4.599) more than Stockfish (4.129 ± 2.624)(p-value < 0.001) (**Figure 1**). Interestingly, the disparity between the two engines' evaluations grew larger as more material was lost.

Space Advantage

Our results show that there was no significant difference between Stockfish's and Komodo's evaluation of the position in regard to space. To represent the balance of total number of moves for either side as a value, the absolute difference between Stockfish's number of advancing moves and Komodo's number of advancing moves (|advW - advB|) were our X-values. The absolute difference between Stockfish's evaluation and Komodo's evaluation (|sEval - kEval|) were our Y-values. We produced no meaningful results by calculating the correlation coefficient for these values, with p-value > 0.1 (**Figure 2**). We also compared X-values to Stockfish's and Komodo's individual evaluations of the position. The insignificant p-values (> 0.1) that we once again produced suggested that the balance in total number of advancing moves for either side does not have a strong correlation with either engine's evaluation of a position.

DISCUSSION

We hypothesized that Komodo and Stockfish would place different levels of importance on material advantage and space advantage. The material advantage portion of this hypothesis was validated by data collected from Chess.com's analysis tab and the use of the Stockfish and Komodo backengines. Overall, the chess engine Komodo valued material advantage more than Stockfish, while space advantage was not greatly valued by either engine. This supports the notion that different engines have different styles of play.

Stockfish has been historically known for its aggressive play. It prioritizes control of the board and weakening the enemy king, while Komodo's play style is defined by its defensiveness and emphasis on pawn structure (13). This fact, coupled with Stockfish's superiority as a chess engine, as indicated by its higher Elo, may serve as a play style guide to chess players, especially in the opening stage of play,

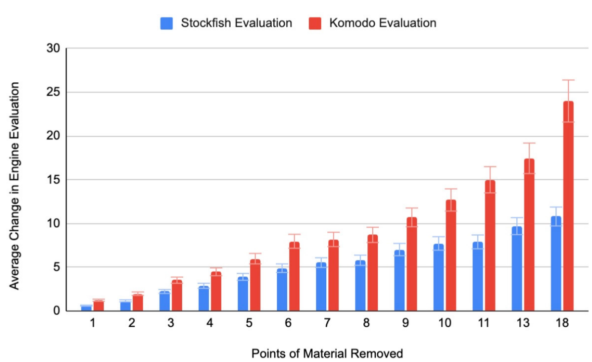

Figure 1: Average engine evaluation change per material removed. A comparison of Komodo's and Stockfish's average change in position evaluation when different points of material were removed. Error bars represent Standard Deviation. Only one test was run with 8, 11, and 18 pieces of material removed. Komodo has a significantly higher sensitivity to material loss than Stockfish. We used a paired t-test, where the change in Stockfish's evaluations and the change in Komodo's evaluations were the samples respectively, and received p < 0.001. The change in position evaluation by both Stockfish and Komodo were each calculated for 80 different iterations over 11 board setups on Chess.com.

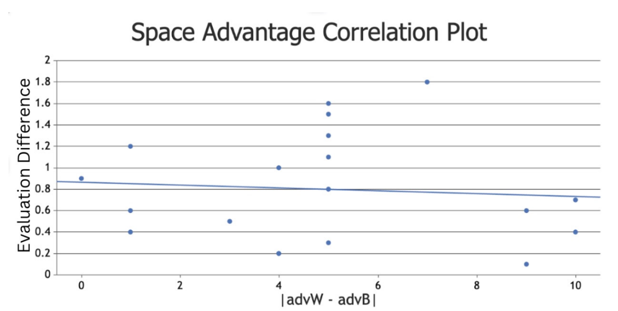

Figure 2: Space advantage impacts on engine evaluation. Displaying the correlation between the difference between Stockfish's and Komodo's evaluations of a position (Evaluation Difference) and the difference between Black's and White's total number of advancing moves (|advW - advB|). Evaluation Difference was calculated by recording Stockfish's and Komodo's evaluation for each chess position generated. |advW - advB| was calculated after counting the total number of advancing moves for black and white for each position. We calculated a Pearson correlation coefficient r = 0.227 between the Evaluation Difference and |advW - advB|, resulting in a p value > 0.1. This indicates that the correlation between the two variables is insignificant.

JOURNAL OF
EMERGING INVESTIGATORS

https://doi.org/10.59720/23-131

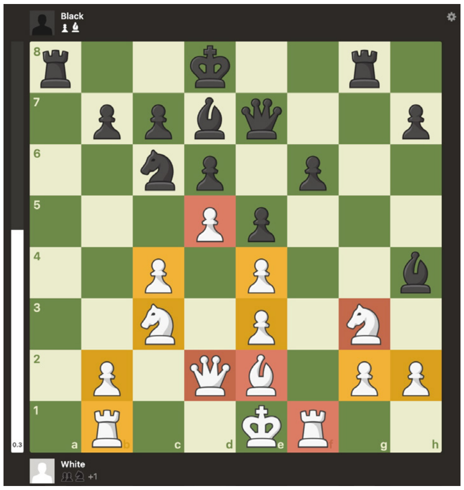

Figure 3: Candidate and non-candidate pieces for removal for material advantage testing. The pieces highlighted in orange represent candidate pieces, or pieces that would potentially be removed for material advantage testing. Pieces highlighted in red, such as the knight in position g3, is not a candidate piece because its removal would open the white king up to check from the black bishop. The d5 pawn is more than halfway up the board from white's perspective, while the pieces on d2, e2, and f1 directly surround the king, and are thus also not candidate pieces.

which sets the stage for the rest of the game (13). Specifically, it may indicate that material advantage should not be given an excessive amount of importance relative to other factors such as space advantage. We speculate that emulation of a chess engine's play style may be beneficial for novice players who have not yet begun to develop their own play style (14) and even for more experienced players who would like to add some structure to the way they play chess. This may also be the case for professional chess players, as chess engines now regularly play well above the level of the strongest human chess players (15). At the highest level, seemingly small differences in a chess position like one side having more piece development or a stronger pawn structure can determine the outcome of the game (15). In-depth analysis of a chess engine like Stockfish or Komodo may well uncover pointers as to how chess players can refine their play style by making small changes.

There exist several popular chess openings that involve the sacrifice of some material to gain an advantage in piece development or control of the central squares of the board. These openings are referred to as gambits and are widely seen as riskier openings due to the consequences that could follow if the aforementioned advantages are not fully capitalized on (16). This is also in line with our findings that losing material is often far more detrimental than losing space advantage.

The data collected for space advantage did not yield sufficient evidence to make a claim about Stockfish's and Komodo's relative use of it for position evaluation. There was no strong correlation in the data collected for space advantage. This could mean that while space advantage is something taken into consideration (17), other factors like material advantage are far more important to engines when evaluating a position.

It should be noted that engines do not think about chess positions in the same way as humans; they determine the degree to which they are winning based on their prior experience in chess and extremely complex algorithms, not based on cut and dry factors like "material advantage" and "space advantage" (18). This, we believe, proved to be the biggest limitation in this study. While testing for certain factors, there were other unaccounted factors that undoubtedly influenced these results. To circumvent this issue, the difference in evaluation was measured before and after piece removal, rather than just using the final or initial evaluation. When testing for Space advantage, the absolute difference between white's number of possible moves and black's number of possible moves was tested for correlation with the absolute difference between Stockfish's evaluation and Komodo's evaluation. We believe that even if more data points were used for space advantage testing, there would be little effect on its significance. As suggested by our results, a larger number of advancing moves, while beneficial, does not directly elevate a player's position.

The limitations of this study led us to ponder the other factors and the degree to which they contribute to an engine's evaluation of a position. We considered whether advancing moves that resulted in the capture of an enemy piece should be weighted more heavily, and if that would produce a relationship between space advantage and engine evaluation. We could determine this more complex relationship between space advantage and engine evaluation of a position through a similar manner to the current space advantage testing, but by increasing a side's advancing move count further for every move that captures an enemy piece.

While our study compared the value an engine placed on material advantage or space advantage with that of the other engine, it may also be possible to relate the value placed on material advantage versus space advantage within the same engine. This could be achieved by conducting the material advantage tests and then recording engine evaluations before and after the number of advancing moves were limited. Then, differences in the engine's evaluation after each test would

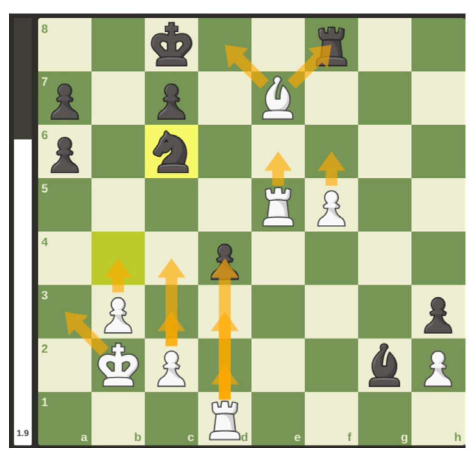

Figure 4: Moves considered to be advancing moves for space advantage testing. Arrows denoting all possible advancing moves each of the white pieces can make in this position. Only legal moves are counted, so the white king moving to position c3, for example, is not considered. This includes piece capture, such as the move Rxd4.

be compared. A drawback of this approach is that limiting the number of advancing moves would involve the movement or deletion of pieces, which would have an effect on other factors that influence the game. Future topics of study could aim to determine what other major influences exist in a chess engine's evaluation.

However, our findings that material advantage is more valued by Stockfish than Komodo and that neither engine greatly values space advantage may help chess players more rapidly improve their skill level by more effectively interacting with chess engines during training.

MATERIALS AND METHODS

Chess Position Generation

To generate the positions, we used chess.com's Explorer resource which contains a database of nearly 3 million games played by its users (19). This resource allowed for the navigation of both commonly and infrequently played chess positions. Using this tool, 11 positions for material advantage testing and 20 positions for space advantage testing were generated, including opening positions (7 - 8 moves played or less), endgame positions (generally characterized by a lack of queens or rooks on the board), and middlegame positions (the remainder of the positions).

Evaluation of Chess Positions

We tested the position evaluation of chess engines on their ability to assess the value of chess positions based on two criteria: material and space advantage. To accomplish this, we utilized chess.com's analysis tool which provides an evaluation of a position based on which chess engine is selected (Stockfish 15.1 or Komodo).

Material Advantage Data Collection

In order to test for material advantage, 11 positions spanning the opening, middlegame, and endgame were chosen. The Stockfish back engine was activated first. After the engine's position evaluation was recorded for each starting position, pieces were randomly selected and removed from one of the sides. The particular pieces that were removed as well as the final evaluation of the position were also recorded.

To conduct the material advantage testing, we began by recording the initial engine evaluation of the position given by Stockfish (E_{Si}). We then selected "candidate pieces" for deletion. These are pieces that are not directly surrounding the king, are not more than halfway up the board, and whose removal does not result in an illegal position. The reason we opted to disqualify pieces more than halfway up the board from being a candidate piece is because we theorized that the removal of these pieces would have a more profound impact on the engine's evaluation due to their controlling squares nearer to the opponent's king. Similarly, pieces directly surrounding the king were not qualified to be candidate pieces as the safety of the king may be another factor chess engines use to determine their evaluation (**Figure 3**).

Pawns, knights, bishops, rooks, and queens were worth 1, 3, 3, 5, and 9 points respectively. The probability of a candidate piece worth n points being removed is 1/(n+1).

Randomly selected subsets of candidate pieces were removed. The same subsets of pieces were removed for Stockfish and Komodo. Once pieces were removed, Stockfish's engine evaluation was recorded again as E_{Sf}. The process was then repeated using Komodo instead of Stockfish, and the initial engine evaluation was recorded as E_{Ki}. The total point loss from one side and the final engine evaluation for Komodo (E_{Kf}) were recorded. In all, data for 80 tests for each engine were recorded across 11 starting positions. The same positions were tested multiple times with different pieces being removed each time, resulting in 160 sets of data points.

Material Advantage Statistical Testing

In order to determine whether there was a significant difference between variables $|E_{Si}-E_{Sf}|$ and $|E_{Ki}-E_{Kf}|$, we used a paired t-test to calculate the p-value comparing the two groups. Our samples for each variable covered each of the 80 iterations of testing.

Space Advantage Data Collection

Setting up a position to test for space advantage followed the same procedure as conducted for materials advantage testing with 20 separate chess positions.

Advancing moves represent how many choices for moves either side would have if it was their turn to play (**Figure 4**). Advancing moves were defined as moves that move a piece further up the board from the playing side's perspective. It should be noted that a rook moving directly sideways, for example, was not considered an advancing move. The reason that only the number of advancing moves for each side were recorded rather than total moves is that backwards moves rarely directly lead to an objective being fulfilled, like the control of squares on the opponent's side of the board or an attack on the enemy king (20).

Stockfish's and Komodo's initial evaluations of the position were separately recorded as well, and we manually counted the number of available advancing moves for white and recorded this value as advW. Flipping the board perspective to black's point of view, the number of available advancing moves for black were recorded as advB. Using Stockfish as the back engine, the position evaluation was recorded as sEval. Then, the back engine was switched to Komodo, and its evaluation was recorded as kEval. Data were collected for each of the 20 chess positions. The two sets of values that were compared were |advW - advB| and |sEval - kEval|. Only absolute differences were recorded because one side having more advancing moves available to them did not necessarily mean that a certain engine would have a higher evaluation than the other.

Space Advantage Statistical Testing

We calculated a Pearson correlation coefficient for the two samples, |advW - advB| and |sEval - kEval|, to find the p-value for space advantage. We opted for correlation due to data being collected from a static position, rather than from a change in position such as in material advantage data collection.

Received: June 12, 2023
Accepted: October 18, 2023
Published: May 14, 2024

REFERENCES

1. Rougetet, Lisa. "Un ordinateur champion du monde d'Échecs : histoire d'un affrontement homme-machine."

https://doi.org/10.59720/23-131

Sciences du Jeu, vol. 5, https://doi.org/10.4000/sdj.598
2. Menon, Aditya. "The impact of Artificial Intelligence (AI) and Engines on Boardgames (Chess and Go)." *International Journal of High School Research*, vol. 4, no.2, https://doi.org/10.36838/v4i2.12
3. Levinson, Robert. "The Case for Computer Chess Engines in the Classroom." *International Journal of Technology, Knowledge and Society*, vol 7, no. 2, https://doi.org/10.18848/1832-3669/CGP/v07i02/56189
4. Fogel, David, et al. "A self-learning evolutionary chess program." *IEEE*, vol. 90, no. 12, pp. 1947 - 1954 https://doi.org/10.1109/JPROC.2004.837633
5. Liaqat, Aisha, et al. "Metamorphic Testing of an Artificially Intelligent Chess Game." *IEEE Access*, vol 8, pp. 174179 – 174190 https://doi.org/10.1109/ACCESS.2020.3024929
6. Tomašev, Nenad, et al. "Reimagining chess with AlphaZero." *ACM Digital Library*, vol. 65, no. 2, pp. 60 – 66 https://doi.org/10.1145/3460349
7. "FIDE Ratings and Statistics." *International Chess Federation*. ratings.fide.com. Accessed 22 Sep. 2023.
8. CCRL. (n.d.). CCRL 40/15. CCRL 40/15 - index. https://computerchess.org.uk/ccrl/4040/
9. Dimitrova, Leyla. "Research of the efficiency of online training sessions among 10-13-year-old chess players." *Journal of Applied Sports Sciences*, Vol. 1, no. 2022, pp. 82 – 88 https://doi.org/10.37393/JASS.2022.01.7
10. Manzo, Andrea and Ciancarini, Paolo. (2023). “Enhancing Stockfish: A Chess Engine Tailored for Training Human Players.” *Entertainment Computing - ICEC 2023*, vol 14455 https://doi.org/10.1007/978-981-99-8248-6_23
11. Gupta, Aditya, et al. "On the Value of Chess Squares." *Entropy*. 2023; vol. 25, no. 10, pp. 1374. https://doi.org/10.3390/e25101374
12. "Chess.com." *Chess.com.* www.chess.com/analysis?tab=analysis. Accessed 22 Sep. 2023.
13. Levene, Mark and Bar-Ilan, Judit. "Comparing Typical Opening Move Choices Made by Humans and Chess Engines." *The Computer Journal*, vol. 50, no. 5, pp. 567 – 573 https://doi.org/10.1093/comjnl/bxm025
14. Cleveland, Alfred. "The Psychology of Chess and of Learning to Play It." *The American Journal of Psychology*, vol. 18, no. 3, 1907, pp. 269-308 https://doi.org/10.2307/1412592
15. Unterrainer, Josef Martin, et al. "Revising Superior Planning Performance in Chess Players: The Impact of Time Restriction and Motivation Aspects." *The American Journal of Psychology*, vol. 124, no. 2, 2011, pp. 213-25 https://doi.org/10.5406/amerjpsyc.124.2.0213
16. Munshi, Jamal. "Evaluation of Chess Gambits." *Social Science Research Network* https://doi.org/10.2139/ssrn.2598867
17. Asatryan, Samvel. "Socio-psychological analysis of factors influencing chess education." *Mankavarzhut'yan ev hogebanut'yan himnakhndirner*, Vol. 20, no. 2 https://doi.org/10.24234/miopap.v8i2.403
18. Kagkas, Dimitrios, et al. "Chess Position Evaluation Using Radial Basis Function Neural Networks." *Complexity*, vol. 2023 https://doi.org/10.1155/2023/7143943
19. "Chess.com." *Chess.com.* www.chess.com/explorer. Accessed 22 Sep. 2023.
20. Maharaj, Shiva, et al. "Chess AI: Competing Paradigms for Machine Intelligence." *Entropy*, Vol. 24, no. 550, pp. 550 https://doi.org/10.3390/e24040550