Finding potential chess cheaters using Spark, SQL, and the Lichess database

Candidate number: 20948

May 8, 2021

1 Abstract

Using Openings Analysis, differences in player ratings; where we see how often the "underdog" wins, and analysing the variation in Openings on a player by player basis, we can create a list of suspected cheaters who are likely using an Engine, sharing the account with other players, or colluding with other players to "throw games" and transfer rating points to a favoured account before closing the dummy account.

2 Introduction

Chess is an ancient game that is thought by many to be out of the reaches of their capabilities due to its complexity. The Ponziani, discovered in literature as early as 1497 is considered to be the oldest recorded chess opening and since then there has been an effort to collate the knowledge about playing the game with the introduction of Openings and Theory [1]. As in real-life, when we are faced with a previously seen situation we use our prior knowledge to make an informed decision, which is what Opening Theory centres on. Bobby Fischer[2] publicly admonished the use of Opening Theory as the reason why he began to hate the game and proposed that Chess was no longer fun (Chess 2020) as no creativity was required anymore if one could just remember lots of Openings and start to play without thinking.

[1] Theory is the current consensus regarding the optimal move to play in each position, compared to a novelty which is a move that has never been recorded in a particular position.

[2] World Champion 1972-175

This notion of "play without thinking" seems very rule based, even robotic. This brings us to the focus of this paper, computer use *during* the game. In the last few decades machines have been created that can estimate what the optimal move in any position is. Combine this with human-learned Opening Theory and we can start to see where Bobby Fischer was coming from. The first big breakthrough came with Deep Blue when it beat the then World Champion Garry Kasparov (IBM 2021). Since then there have been incremental improvements until Deep Mind released AlphaZero in 2017 which uses Machine Learning and Reinforcement Learning to make more "human-like" decisions (Silver 2018).

The introduction of chess engines is argued by most to be beneficial as it allows players to improve their play and hopefully lead to more interesting games. There is no rule that says a person can not use an engine to train, but they are expressly forbidden during games. This is not a problem during Over The Board (OTB) games which are played in person, but this is a problem in online games where it is almost impossible to check a person is not using an engine to estimate the best reply to their opponents move.

This type of cheating has been a problem since the introduction of engines but as they become more accessible, and often free, it is argued that cheating has increased. Using the database from Lichess.org I will conduct research into whether we can create a list of potential cheaters that can then be further investigated (Lichess 2021). I can not conclusively say they were cheating, but given my knowledge of the game after over 3,000 games I feel I am aware of some of the traits of cheaters and hopefully the data will shed light on some of these (Chess.com 2021). I will discuss in the Extensions section the steps needed to further check if these candidates have indeed cheated.

With the explosion in popularity of chess due to the Netflix series "The Queens' Gambit" it has never been more important to make sure that online chess retains a reputation of fairness and being in the main, free from cheaters. With this new influx of people the burden on the people who decide whether, given the evidence that a person is cheating, has increased dramatically. Currently each player is assessed on a player by player basis where a moderator who works for the platform will look at evidence submitted by the rest of the community[3].

Therefore, the need for this process to be automated has never been greater. I hope my research can lay the foundations for achieving some of

[3]Usually in the form of specific games that are believed to be "unusual"

this and alleviating the initial checks that the moderators must do[4].

3 Data

I had problems conducting the analysis on the larger data sets as discussed more thoroughly in the accompanying Jupyter Notebook, so here I will use the January 2013 dataset which contains 121,332 games. Even though I am using a small data set, the same functions and techniques, Spark, SQL, RDDs, etc are immediately transferable to larger data sets so I believe the skills learnt in creating this paper are not wasted.

Even though I am only using the January 2013 game data, it still contains games that span the full range of player ability from the very beginner to Grandmaster level. An interesting effect of using older data is we can directly check the Lichess.org website to see if any of our candidate cheaters infact did have their accounts closed.

For each game we have the following features:

1. Event: The name of the type of game, such as Rapid, Daily, Classical, etc. These are categories of games that can include multiple Time Controls, but generally have similar characteristics as the maximum time a player has for the game can entice different types of players - some prefer slower games, others very fast.

2. White & Black: Usernames of the respective players in the particular game.

3. Result: Either "1-0" indicating White won, "1/2-1/2" indicating a draw, or "0-1" indicating Black won.

4. WhiteElo & BlackElo: Ratings of each player. As described below ELO, Glicko 1, Glicko 2 do differ, but generally only by a constant so whilst I will aim to refer to Glicko 2 as per the Lichess documentation, we can use these names somewhat interchangeably.

5. Opening: A text version of the Opening played.

[4]This objective filtering should also improve the reputation of the platform as a concrete list of what moderators are looking for when they decide someone may be a candidate cheater can be released - rather than currently where it is often based on intuition that has been fostered over many years of playing chess.

6. TimeControl: The number of seconds per player to make all of their moves. If either player's time expires then they forfeit the game, irrespective of if they had the better position.
7. Termination: A description of why the game ended. "Normal" refers to one player achieving checkmate, "Time Forfeit" denotes that one player lost because their TimeControl expired.

3.1 Glicko 2 Rating

Chess organisations and federations use the ELO rating system to estimate a player's ability. Lichess uses the Glicko 2 rating (GR) system which is argued to be more rigorous and takes account for the volatility in a player's recent results. The rating system used for the other popular chess website (Chess.com) is the Glicko 1 rating system. This means we are unable to merge our data sets and must stick to one. Ratings names are often used interchangeably and even though they do differ, can often result in similar trajectories, albeit shifted by some constant. Therefore, the analysis we will do should lead to similar results for other rating systems as the underlying signs of a person cheating are the same.

3.1.1 Rating Binning

Even though ratings are integer values, and range from around 500-3000, and are hence finite, the difference between similar ratings is not noticeable, for example there is no feature that makes a 700 rated player vastly different than a 701, 702, etc rated player. This only occurs when the gap becomes sufficiently large, e.g. a 700 player will likely play slightly differently than an 800 rated player. I will group similar ratings together into bins of 15 equally sized bins between the minimum and maximum GR of the data set.

4 Methods

I will attempt to create a candidate list of cheaters/Engine users by analysing the Openings used by the players in Subsection 4.1. After which we will look at the variation in Openings. It is common for Grandmasters to know many Openings in great detail, but by the nature of being a beginner and not having the knowledge to get them to Grandmaster level, we would expect other players to only know a few openings. Therefore, it is suspicious if a player plays many different Openings, and especially so if they win with a lot of different Openings. In Subsection 4.2 we will look at the players with the most varied Opening repertoire. The rules of chess state that White always plays first.

Then we will look at the differential in GR between players to see which players beat players who were much high ranked than them in Subsection 4.3. A low ranked player may occasionally beat much higher ranked players, but if this is happening consistently then it may be a sign of Engine use.

It is this quirk in the game that means that White has the initial Tempo and gets to control the game at the start, and hence which direction the Opening will take. As we will see, it should be more apparent if a player is candidate cheater if when they are Black they have a large Openings repertoire. Hence Subsection 4.2's analysis will be done from the perspective of Black, whereas Subsections 4.3 and 4.1 are completed from the perspective of White.

4.1 Openings Analysis

There are many thousands of Openings and this can be overwhelming for a new player. Therefore, it is common for beginners to start with some canonical Openings such as The London System for White, and the Caro Kann or Queen's Gambit Declined for Black. It is therefore, common for a lot of beginner level games to incorporate these. However, they are not necessarily the best Openings[5] but are those that are surmountable for beginners. Grandmasters are likely to favour other Openings that may prove trickier to play against, but do require more studying to play them well and not make fatal mistakes. Grandmasters will also likely avoid beginner Openings as they are aware that their equally high rated opponent knows how to defend against them, and hence playing the Opening is trivial and would likely lead to a draw. It is unlikely that we would see a low level player playing Openings that are considered too difficult for their GR bin[6]. Alone this is not a definitive sign of a candidate cheater, but does give cause for concern.

Whilst we would not expect a lot of Grandmasters to play beginner Openings, we would expect that when they do, that they either win or draw. White

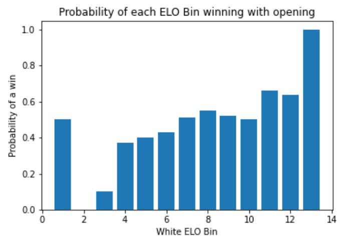
[5]The notion of any set of best Openings is often contested anyway

[6]Whilst it is naturally not impossible for this to happen, by necessity, if they were legitimately playing these Openings then their GR would rise and they would become a Grandmaster.

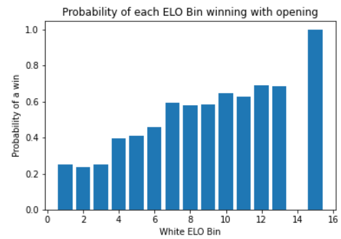
has the advantage of always going first so it is up to Black to force White to make a mistake and hence create a weakness. If White is knowledgeable about the opening then it is more likely that they will win, or if Black can defend well, draw. Therefore, we should expect as GR increases that the probability of a win for White or a draw increases for the beginner openings.

It is possible that even though a beginner plays a Grandmaster Opening, that they may still win legitimately as it may be that the opponent blundered a piece, irrespective of the Opening used. This is to say that not every beginner who plays a Grandmaster Opening and wins is necessarily a cheater. Games are by default created between two similarly ranked players. However, during tournaments this is not as stringent a rule and the gap between two players may increase more. With the additional incentive to win online tournaments we could expect the motivation to cheat to increase. If White is a much lower rated player, but cheating and using a Grandmaster Opening, but still beats a much higher rated player, then we can postulate that the win was in this case more likely down to something that White is doing, rather than rudimentary Blunders[7] that Black is making.

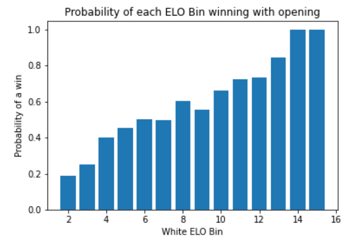
[7] A critically bad move that can often cause a player to lose the game, or at minimum a lot of the advantage they may have had.



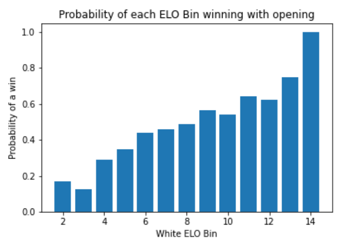
(a) French Defense: Knight Variation



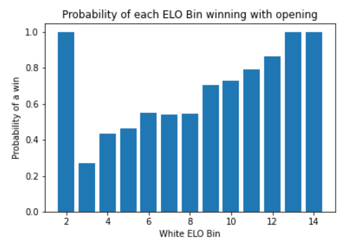
(b) Horwitz Defense



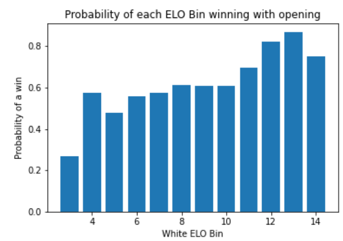
(c) Modern Defense



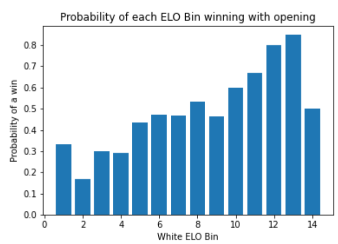
(d) Owen Defense



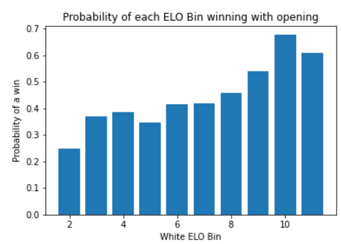
(e) Philidor Defense 3



(f) Scandinavian Defense 1



(g) Scandinavian Defense 2



(h) Sicilian Defense

Figure 1: Top White Openings and the probability of a win for each ratings bin

We can see in Figure 1 that Openings such as the ”Owen Defense” and the ”Modern Defense” have much higher win ratios for the upper ratings bins and I would class these as Grandmaster Openings, as only the Grandmasters were able to win with them. Openings such as the ”Sicilian Defense” and the ”Scandanavian Defense 1” are more evenly distributed across all ratings bins. This suggests it is an Opening that all levels of players can play.

An extension to this paper would be to find low rated players that are playing Openings such as the ”Owen Defense” and winning a high proportion of games. In general we have seen that low rated players do poorly with this Opening so any low rated player who wins excessively is a candidate cheater.

4.2 Scope of Openings repertoire

Grandmasters have the ability to play many Openings well, even though they are still likely to have their favourites that they play substantially more than other Openings. However, a beginner will have a much narrower Openings knowledge and hence we would expect a legitimate player to stick to a few Openings. Legitimate players can trial out new Openings though to see if they prefer the style of a new one compared to one they currently player, so the mere fact they played many Openings is not sufficient to say a person is cheating. What is more suspicious is if they play many Openings, and have a good win record with this vast array of Openings.

Except for a few edge cases such as the Caro-Kann Defense, it is usually White who forces the hand of Black and because White always plays first. Therefore, even if the player was a cheat, if they are playing with White then they have no need to play many different Openings, and likely draw attention to themselves, their best bet is to use the same Opening - with an Engine, and force Black to respond. = Therefore, we will analyse the data from Black's perspective for this section. If Black can't control White's first move then it will need a larger repertoire of responses. This is still true for a legitimate Black player, but even if the response the legitimate Black player likes to play isn't ideal for White's move, they may still play it anyway as they are comfortable with it. An Engine, however, would recommend playing something optimal.

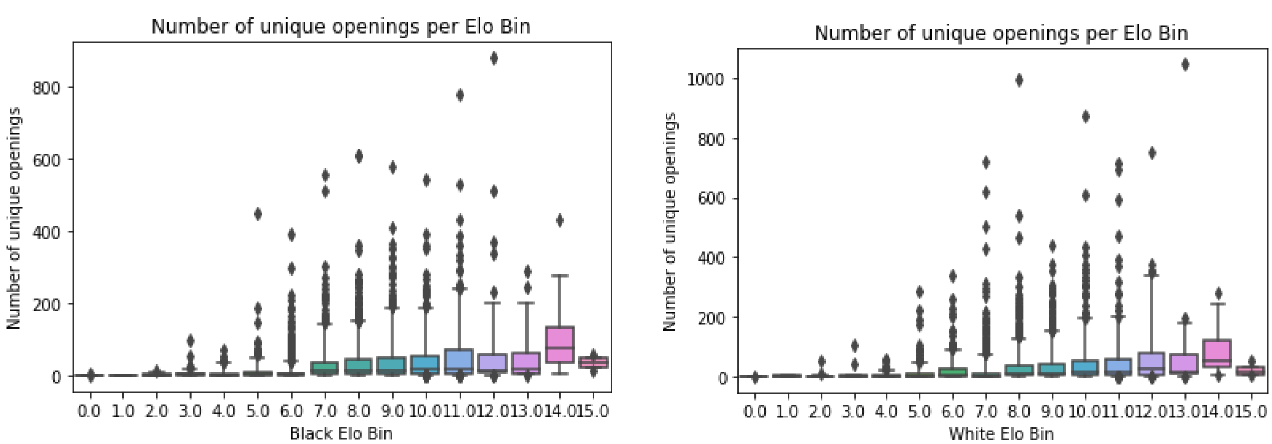


Figure 2: Boxplots showing the distribution of the number of unique Openings played for each rating bin. These were created with the *plot_unique_openings* function.

We can see that the lower rating bins do indeed stay with only playing a few distinct Openings per player and as the players' ratings increase, they each start to play many more Openings. Near the upper end the Grandmasters revert back to only playing a handful of different Openings each. This is in line with our theory that Grandmasters prefer to play a few Openings very well, rather than multiple Openings at a lower quality.

The boxplots show a handful of players that are far above the 95th percentile and play many more Openings than expected. Whilst this is unusually generally, it is less suspicious when it occurs for the White player as they have control because they always play first. It becomes more suspicious for the Black player as they are having to react to potentially unseen positions. If you were a legitimate player you would revert to optimum theoretical play[8], or your known Openings. A cheater may react with a new unique Opening that is optimal, but naturally comes from the Engine.

Indeed we do see a more pronounced, and higher, bar on the 14th class for Black versus White. Unusually though the extremes of classes 8, 10, and 14 are much higher for White than Black. This goes against my intuition but we will see if any of those players fail our other tests before making a judgement.

[8] Such as trying to control the centre of the board, not putting knights on the edge of the board, avoiding blocking your bishops' path

White	Count	Black	Count
Panevis	352	TrialB	350
sport	312	luciosergio	308
sokol88	303	argo	141
WAHAHAHA	178	Kiriush	124
luciosergio	158	1130	81
rashit59	131	rashit49	63
MihaSAH	124	NewAtChess	62
legend999	68	Luminosity	51
Kiriush	68	pseudoknight	44
pseudoknight	62	sakrat	33
TrialB	59	JohnUA	33
crystalcat	54	DiamanteNegro75	31
sakrat	47	Yarilo	25
kramer	35	all-others	24
esmir	22	S_M_D	22
JohnUA	19	URIAHURIAH	15
AL3103	13	pakap	15
1130	11	LLUNA	10
Malkav	6	sokol88	6
miooim	1	serxi	2

Table 1: Number of unique Openings played per player over and above the 95th percentile of their respective ratings bins.

Many of these players appeared on both lists, for Black and White. I think this is suspicious. This suggests that, if legitimate, that these players known - and use - vastly more openings than 95% of all other players, for both Black *and* White, which, to me, seems unlikely.

These numbers represent the number of unique Openings that the player played above the 95th percentile, for their respective rating bin.

4.3 Differentials between player ratings

Function: player_differential_white and *player_differential_black*

I analysed the win ratio between pairs of players where the differences in ratings was above a certain threshold. I did this using Spark SQL whereby

I selected all of the games that resulted in a win for the respective colour, and then by games where the difference in ratings was above the threshold. This was done from each colour's perspective where I looked at the games where White lost, but was much higher rated, and vice versa from Black's perspective. I then grouped all of those games by player to see which player beat much higher rated opponents most often, which may suggest Engine use.

White	**Count**	**Black**	**Count**
miooim	12	NewAtChess	23
JohnUA	12	kramer	18
FCST1923	10	khorzokhan	11
kramer	10	Karen_Armenia	11
barriosgb2	9	luciosergio	9
khorzokhan	9	JohnUA	8
Joueurfel	9	miooim	8
argo	8	Joueurfel	8
WealthandTaste	7	twochief	7
fil77	7	barriosgb2	7
bsn32	6	zer001	7
migsan10	6	argo	7
jansom	6	Mafia_top	7
-King	5	DaveBloom	6
Robocop	5	Caboose	6
TrialB	5	rastikk	6
SAIRAM	5	bsn232	6
khan	5	TrialB	6
DaveBloom	5	karpov999	6
sokol88	5	tarekegypt	5

Table 2: Count for the times a player has beaten a player who is 300 or more rating points higher than they are.

10 of the 20 players appear on both lists. I believe this is suspicious. Players such as "kramer" and "khan" have indeed been banned which suggests that this method of looking at how often a player, who is much lower rated, wins, is a successful filter for cheaters. 300 rating points is something that

would take an average player who is playing an hour a day, many months to achieve so I believe beating a player who is 300 points higher is very unusual[9]

5 Extensions

A common "tell" if a person is using an Engine is a consistent time to play each move, regardless of the move. To cheat using an Engine the player must enter the opponents move into the Engine, wait for it to calculate the optimal move, the player must then interpret this, and finally make the move on the actual game board. The Engine is often very consistent with how long it takes to find the optimal move. This leads to a consistent pattern in the time taken for each move during the game. This alone might be an indication of cheating. However, when we consider whether this happens on obvious moves, or moves that require more thought it becomes apparent that a legitimate player is more inconsistent as they consider their options. As most players use Opening Theory today the first 10 moves of the game are usually in a player's memory and any variations depending on the opponent's move are also memorised so these moves can be played often without hesitation, compared to more difficult moves later in the game that might require a human to think more. Therefore, legitimate games often have varying times per move. I did not have access to the time per move data for each game but I think this is an important extension to this paper in the future.

I would have liked to create a script that checks the Lichess.org website to check if our candidate cheaters did indeed have their account closed, which would allow us to calculate an accuracy score.

6 Limitations

Some accounts fail all of our tests and look like obvious cheaters. However, they are just very prolific players, some with tens of thousands of games. This makes it plausible that they do indeed just play a lot of openings, and because they play so often, can indeed beat much higher rated players by the fact that they have more opportunities to do so. This means that whilst our analysis reduces the list of potential cheats substantially, it still requires some human intervention to decide if these candidates need to have their accounts closed.

[9] For a point of reference, I have only ever beaten a player who is 40 rating points above me.

Some players will have moved from one rating bin to another and hence on some analyses will appear twice. This did not become a problem for the *top 20* tables I created but might be a slight over estimation if we were interested in the total count of players failing any particular test.

7 Conclusion

Some players did indeed fail multiple tests as can be seen by Table 1, and 2. Pleasingly a lot of these players have since been banned, such as ”khan”, ”kramer” and ”NewAtChess”, of which our model predicted those players as being cheats.

There is still work to be done and incorporating move times should improve the analysis but for a proof-of-concept I believe this paper fulfills its aim to create a candidate list of cheaters.

References

Silver, David (2018). "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". **in**: *Science* 362.6419, **pages** 1140–1144.

Chess, Dark Knight (**august** 2020). **in**: *YouTube*. URL: `https://www.youtube.com/watch?v=P349BdHUxlc`.

Chess.com (2021). *OrangeEcon - Chess Profile*. URL: `https://www.chess.com/member/orangeecon`.

IBM (2021). "Deep Blue". **in**: *IBM100 Deep Blue*. URL: `https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/`.

Lichess (2021). URL: `https://database.lichess.org/`.