Analysis of Lichess Data

Andrew Mezzi

Contents

Overview of the Data

This data is based on 18,000 chess games played on the popular online chess website Lichess. The following statistics were collected for each game:

Elo Rating

Elo rating (named after its inventor Arpád Élő) is a family of methods for rating players in two-player games. It is the most common rating system in in chess. It works as follows. Imagine there are two chess players, Alice and Bob. Alice and Bob are both given an expected score[1] based on their Elo that represents their probability of winning: an expected score of 1 represents a 100% chance of winning, an expected score of 0 represents a 0% chance of winning, an expected score of 0.75 represents a 75% chance of winning, and so on.

A chess game has three possible outcomes: a win for Alice (Alice 1–Bob 0), a win for Bob (Alice 0–Bob 1), or a draw (Alice 1/2–Bob 1/2). After the game, Alice and Bob's expected scores are compared to their real scores. If Alice's score is higher than her expected score then her Elo was too high and should be decreased while if her score is lower than her expected score, her Elo was too low and should be increased. Taking this into account, Alice and Bob's Elo scores are adjusted to fix this.[2]

In practice, if a player wins a match against a player with a higher rating than them, their rating will increase a lot. If a player wins against a player with a lower rating than them, their rating will only increase a little. If a player loses against a player with a lower rating, their rating will decrease a lot. If a player loses against a player with a higher rating, their rating will decrease a little.

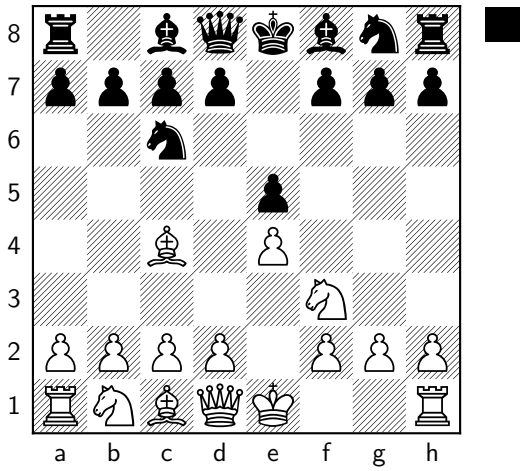
Opening ECO

The first few moves in a chess game, called the *opening,* have been studied extensively over the years and almost every possible sequence of moves has been named and classified. The standard way to refer to openings is using their *Encyclopaedia of Chess Openings* code. For example, the Italian Game (1.e4 e5 2.Nf3 Nc6 3.Bc4), often the first opening that new chess players learn, has the ECO code C50.

1 e4 e5 2 ♘f3 ♘c6 3 ♗c4

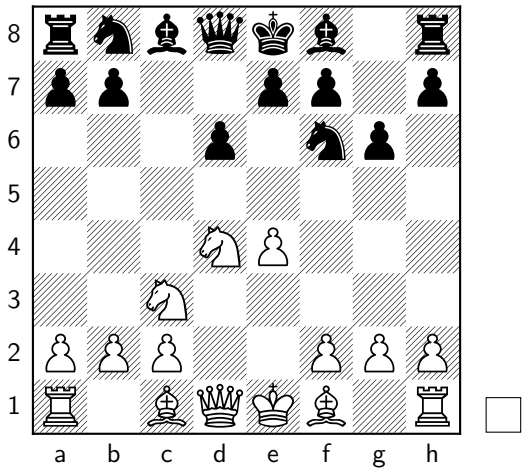
[1] $E_A = \frac{1}{1+10^{(R_B - R_A)/400}}$ where E_A is Alice's expected score, R_A is Alice's Elo rating, and R_B is Bob's Elo rating.

[2] $R'_A = R_A + K(S_A - E_A)$ where R'_A is Alice's new Elo rating, R_A is Alice's old Elo rating, K is the maximum possible change in the rating per game (usually eith 16 or 32), S_A is Alice's score, and E_A is Alice's expected score.



Meanwhile, the Sicilian Dragon (1.e4 c5 2.Nf3 d6 3.d4 cxd4 4.Nxd4 Nf6 5.Nc3 g6), which is a very popular opening in upper-mid level play, has the ECO code B70.

1 e4 c5 2 ♘f3 d6 3 d4 c×d4 4 ♘×d4 ♘f6 5 ♘c3 g6



Opening Ply

While the first few moves have been studied extensiely (in chess lingo, these are called "book moves"), every chess game will eventually reach a new position ("moving out of the book") where players will need to rely on their tactical skill instead of memorization. The Opening Ply gives the number of moves until players move out of the book.

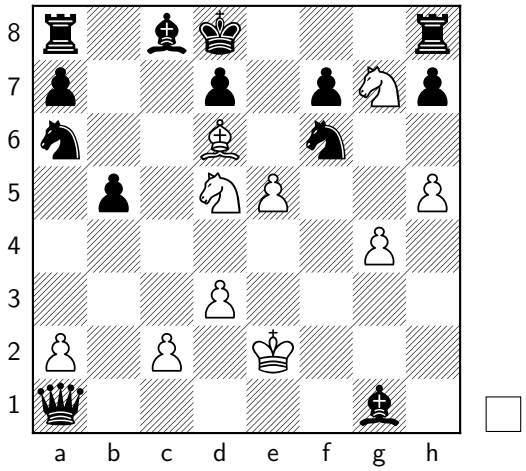
Centi-pawn loss

Some chess pieces are more valuable than others. For example, a queen is obviously more valuable than a pawn. Chess pieces are valued in centi-pawns: traditionally, a pawn is worth 100 centi-pawns, a knight or bishop is worth 300 centi-pawns or 3 pawns, a rook is worth 500 centi-pawns or 5 pawns, and a queen is worth 900 centi-pawns or 9 pawns.[3] This means that sacrificing three pawns for your opponent's bishop, or two pawns and a knight for your opponent's rook, is (usually) an equal trade.

[3]These numbers aren't exactly correct, rather they're estimates that are easier for players to calculate with during a game. According to the leading chess engine AlphaZero, a knight is actually worth 305 centi-pawns, a bishop is worth 333 centi-pawns, a rook is worth 563 centi-pawns, and a queen is worth 950 centi-pawns

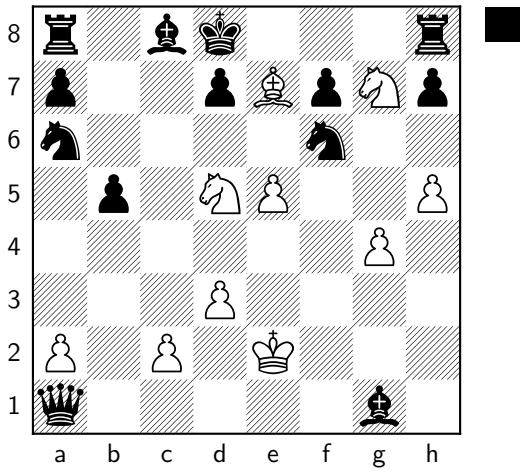
However, chess pieces are worthless in a vacuum. Just because Alice has better pieces than Bob doesn't mean that Alice will win. Consider this famous game played between Adolf Anderssen (as white) and Lionel Kieseritzky (as black) in 1851.

1 e4 e5 2 f4 e×f4 3 ♗c4 ♕h4+ 4 ♔f1 b5 5 ♗×b5 ♘f6 6 ♘f3 ♕h6 7 d3 ♘h5 8 ♘h4 ♕g5 9 ♘f5 c6 10 g4 ♘f6 11 ♖g1 c×b5 12 h4 ♕g6 13 h5 ♕g5 14 ♕f3 ♘g8 15 ♗×f4 ♕f6 16 ♘c3 ♗c5 17 ♘d5 ♕×b2 18 ♗d6 ♗×g1 19 e5 ♕×a1+ 20 ♔e2 ♘a6 21 ♘×g7+ ♔d8 22 ♕f6+ ♘×f6



If we only add up the pieces on the board, Anderssen's are worth 12 pawns while Kieseritzky's are worth 27. Clearly Kieseritzky has the advantage, right? Not so fast. The next turn, Anderssen makes the move Be7, putting Kieseritzky in checkmate and winning the game.

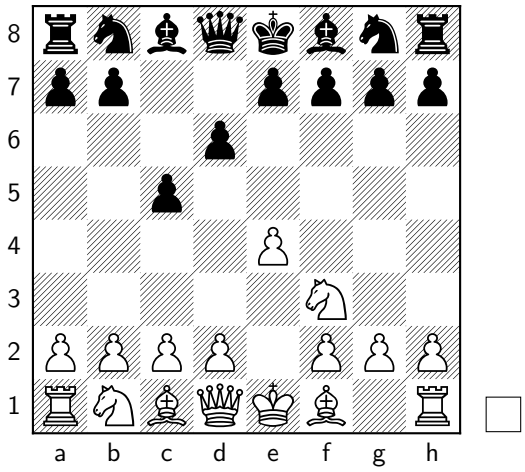
1 e4 e5 2 f4 e×f4 3 ♗c4 ♕h4+ 4 ♔f1 b5 5 ♗×b5 ♘f6 6 ♘f3 ♕h6 7 d3 ♘h5 8 ♘h4 ♕g5 9 ♘f5 c6 10 g4 ♘f6 11 ♖g1 c×b5 12 h4 ♕g6 13 h5 ♕g5 14 ♕f3 ♘g8 15 ♗×f4 ♕f6 16 ♘c3 ♗c5 17 ♘d5 ♕×b2 18 ♗d6 ♗×g1 19 e5 ♕×a1+ 20 ♔e2 ♘a6 21 ♘×g7+ ♔d8 22 ♕f6+ ♘×f6 23 ♗e7#



Instead of counting pieces, chess engines quantify entire board positions in centi-pawns based on how many pawns ahead ahead white (or black) will be if they play perfectly.[4] Consider a mainline Sicilian game (ECO code B50):

1 e4 c5 2 ♘f3 d6

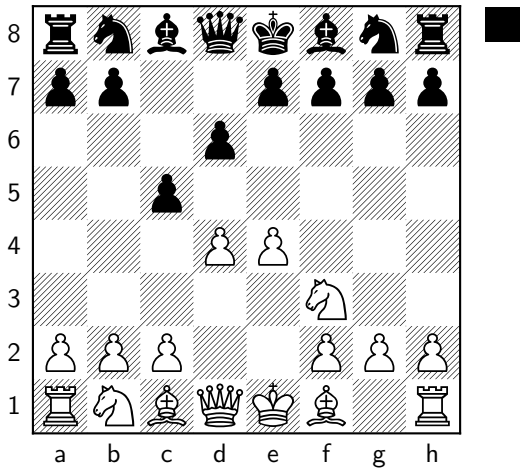
[4]Engines usually do this by calculating every possible combination of moves for the next few turns, seeing how many pawns ahead white is in each game, and weighting this by how likely black is to make those moves.



According to the Stockfish engine (calculating 15 turns ahead), white is 66 centi-pawns ahead of black. In other words, if white moves perfectly for the next 15 turns, on average they will end up with 66 centi-pawns more pieces than black.

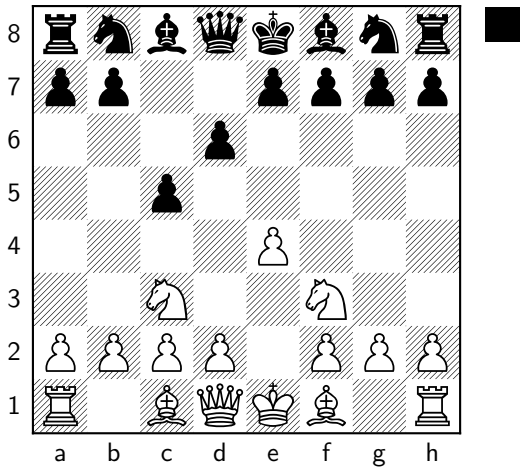
The best possible move for white is d4.

1 e4 c5 2 ♘f3 d6 3 d4



White might not make the best possible move, though. Say white moves Nc3 instead.

1 e4 c5 2 ♘f3 d6 3 ♘c3



This is still a good move, but it's not as good as d4. Stockfish estimates that white will be only 51 centi-pawns ahead with this move. Since this move puts white 51 centi-pawns ahead while the best possible move puts white 66 centi-pawns ahead, white loses 15 centi-pawns by making this move.

The average centi-pawn loss represents how many centi-pawns a player loses on average per turn, compared to the best possible move they could have made.

Number of Blunders, Number of Mistakes, Number of Inaccuracies

These categories are self-explanatory. A blunder is a move that's disastrous for the player who makes it if their opponent exploits it correctly, a mistake is a move that's bad but not necessarily fatal, and an inaccuracy is a move that is decent but worse than the best possible move.