

Choosing chess openings for TCEC10

Nelson Hernandez *
Maryland, USA

TCEC is the Top Chess Engine Championship (TCEC, 2018). Better known by my TCEC nickname of Cato the Younger, I was the person in charge of the opening positions for TCEC Season 10 (Haworth and Hernandez, 2018). I have been doing this since Season 5, first for this tournament's founder Martin Thoresen, and more recently for Anton Mihailov, our tournament director and the owner of chess website Chessdom.com.

For this season, I selected the opening positions for the first two stages while Superfinal positions were selected by my friend and associate Jeroen Noomen. Jeroen has been involved with computer chess for over 35 years and is one of the foremost individuals in the world in the area of computer chess openings. My qualifications rest essentially on one thing, my proprietary database. I've been collecting quality chess games for 13 years and I add millions of new games every year from many sources.

In Stage 1, the 24 contestants faced each other once in a single round-robin event. In other words, every program played 23 games. At the end of Stage 1, the leading eight advanced to Stage 2.

One of our conventions at TCEC is that we start our games from a wide variety of positions in all stages. Why do we do that? There are two reasons.

Firstly, because we think that a chess program, to earn the title of champion, should demonstrate its superiority in a full range of chess positions, not just the ones it likes best. Secondly, starting all games from the traditional opening position results in many programs – not all, but many – playing the same moves at the start of the game, which results in repetitive chess. We tried running a Stage with the opening position a few seasons ago and saw an awful lot of French Exchange openings. The consensus was that we should go back to varied openings.

Now, how you implement varied openings is no trivial matter. There are several considerations to take into account.

You want to represent many different opening systems. You want to have a desirable play-balance as indicated by the information you have at hand, namely your statistics and the evaluations of leading engines. In other words, it won't do if a position is totally one-sided or totally drawn. You don't want positions that have already exchanged off a lot of material or are about to. We prefer positions that have been seen often in our database so that our statistics can help us judge whether a position is one-sided or drawish.

You don't want to follow opening theory too deep into the game because that excessively limits an engine's freedom of choice. You don't want to start from a position where there is really only one conceivable move, because in a case like that you are robbing the engine making the first move of calculation time it might profitably use later in the game.

*E-mail: nhernandez@hotmail.com.

1 Finally, taking everything into account, we want to be fair to every seeded program. We don't want
2 any program to have a material structural disadvantage that is a result of anything we did.

3 Now, this last point was most important for Stage 1 because Stage 1 was a single round-robin format
4 with an even number of programs competing, 24 programs, to be exact. And we're using over 100
5 different two-move openings over the course of Stage 1. There are two problems with this.

6 The first is that all two-move openings are not equal. Statistically and in terms of starting evaluations,
7 they vary considerably.

8 The second issue is that, with 24 engines playing 23 games each, half of the engines will play white
9 12 times and black 11 times and half vice-versa. Naturally that gives the engines playing white 12
10 times a slight advantage, all other things being equal.

11 Given these facts we could not just randomly assign openings to the 276 games of Stage 1. If we had
12 done so, some programs would inevitably have been considerably advantaged or disadvantaged. We
13 might not be able to quantify it exactly, but that unfairness would be there and it would affect every
14 program to varying degrees. Just imagine how bad it could be if one particular engine had got nothing
15 but weak openings to play.

16 I think this was by far the hardest challenge to constructively address in this playing format. But in
17 the interests of fairness we had to make some effort to level the playing field. Otherwise, we would
18 have needed to go to a double round-robin which would have been impractical in this format.

19 So how did I address the problem? I could discuss the fine details of this topic at length: instead, let
20 me cover it in broad terms.

21 There are presently 368 common two-move openings in my library. Each of the resulting starting
22 positions is scored for its favorability to White based on empirical and evaluative data. For Stage 1's
23 set of 276 games I ran over 100,000 trials on an Excel macro that I developed until I got a solution
24 that passed three different criteria for each of the 24 programs.

25 What were these criteria? For each of the 24 programs, all white openings had to be different and all
26 black openings had to be different. In addition, all 24 programs had to fall within a tight fault tolerance
27 in terms of fairness for the entire 23-game opening set.

28 Were the opening sets for all 24 programs perfectly balanced? No, I'm afraid not. There can be no
29 perfection in single round-robin tournaments with different openings. But with a lot of effort you
30 can minimize opening variances to a remarkable degree, resulting in what I would deem to be a fair
31 tournament.

32 As a further security measure, I never knew in advance how the programs would be seeded, and our
33 tournament director never saw my calculations. So before we started the tournament, nobody knew
34 which engines would have a tiny advantage or disadvantage in Stage 1.

35 For Stage 2, each of the eight programs played its seven competitors four times, 28 rounds altogether.
36 Each program played both sides of each position against each opponent. Altogether we saw 56 dif-
37 ferent opening positions. Unlike previous seasons where I used eight-move openings exclusively, this
38 season my openings ranged anywhere from three to twelve moves.

39 I have already covered the factors we took into consideration in Stage 2 and the Superfinal, but I
40 do want to add a word about our opening picks and their relative play-balance, especially to TCEC
41 newcomers.

42 Our openings were deliberately skewed to one color or the other in order to increase the potential for
43 a decisive result. We didn't want to increase that potential so much that we wound up with a one-
44



Fig. 1. Nelson Hernandez, giving his talk on TCEC openings.

sided game. That would have been the worst result from our perspective because it would do nothing to clarify which program was stronger and it would waste everyone's time. Two-sided draws aren't helpful either but at least we could create positions where there was a chance that they wouldn't happen. We're always happy to see a position which is successively won and drawn by a program because in effect it 'won' that opening and thus it offers us a small measure of truth.

Thanks to Anton Mihailov of Chessdom.com for putting on this, the Internet's premier computer chess tournament. I hope you enjoyed Season 10 as it happened and enjoy revisiting the most notable games.

REFERENCES

Haworth, G.M^c. & Hernandez, N. (2018). TCEC10: The 10th top chess engine championship. *ICGA Journal*, 40(2), 113–118. <http://centaur.reading.ac.uk/75887/> for supporting data – results, statistics and pgn files. doi:10.3233/ICG-180045.

TCEC (2018). <http://tcec.chessdom.com>. Current and past TCEC tournaments.