Computationally Estimating the Solvability of Forced Mate Sequences

Azlan Iqbal[1]

ABSTRACT

In automatic chess problem composition, there may be a need to classify or label problems in terms of solvability (i.e. how difficult it is for the typical player to solve) for more efficient human consumption. Automatically classifying a chess problem as being easy, moderate or difficult therefore helps in that regard. In this article, we explain a formula for this that relies on five variables, i.e. chess engine solving time, total piece count, total piece value, move length and an adjusted number to represent the variations possible. The method was tested on samples of forced mates (3, 4 and 5-movers) including both published chess problems by human composers and computer-generated ones. Statistical analysis of the results was significant and as expected based on the groups tested. Such problems may therefore then be classified as 'easy', 'moderate' or 'difficult' based on percentile. The lowest third of scores is considered easy, for example. Longer sequences that do not end in checkmate (such as studies) is still an open problem in terms of solvability using this approach mainly because even good chess engines often cannot solve them conclusively enough to establish the time variable and the number of variations possible.

1 INTRODUCTION

Chess problems or puzzles found online and published in books usually present potential solvers with a position and stipulation such as "*White to Play and Mate in 3 Moves*" or "*White to Play and Draw*". The positions may not necessarily have been composed and could even have been taken from real games. The Wikipedia (2018a) page on this topic provides some insight into a finer distinction by stating the following.

While a chess puzzle is any puzzle involving aspects of chess, a chess problem is an orthodox puzzle in which one must play and win or draw a game, starting with a certain composition of pieces on the chess board, and playing within the standard rules of chess.

The Wikipedia (2018b) page on 'chess problem' further states the following.

The term "chess problem" is not sharply defined: there is no clear demarcation between chess compositions on the one hand and puzzles or tactical exercises on the other.

The terms 'chess problem' and 'chess puzzle' are therefore used interchangeably in this article to mean orthodox puzzles and also include 'chess constructs' (Iqbal, 2014), a typically simpler form of the traditional chess problem. Chess problems are often classified in terms of difficulty so that solvers are better prepared when faced with them. This is so that they are able to make more suitable choices relative to their playing or solving ability. Online chess communities may use a community ranking feature where players vote on how difficult a problem or puzzle was to them and with sufficient participation and time, a realistic estimate of the difficulty of a problem can be established. Others may rely on chess experts, masters or trainers to provide an estimated classification based on their experience. Some may even have their own Elo-like rating systems for problem solving, even though these would be difficult to validate outside of controlled conditions (such as a problem solving contest) where solvers did not have access to chess engines. GameKnot (2018), for instance, presents puzzles as shown in Fig. 1 and has the following to say.

The goal of all chess puzzles is to checkmate your virtual opponent no matter what moves they make (i.e. a forced mate), in the requested number of moves. Some chess puzzles are created from actual chess games played online, and some are purely composed chess problems, sometimes even with positions that cannot be reached in a real game of chess. All chess puzzles are automatically verified, so all solutions are guaranteed to be correct and complete. However, there is always a chance that a shorter solution exists, so if you believe you found a better solution to any of the chess puzzles, please use Options → Another solution in the bottom right corner of the puzzle window to enter it.

[1] College of Computing & Informatics, Universiti Tenaga Nasional, Kampus Putrajaya, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia. E-mail: azlan@uniten.edu.my

Figure 1: Chess puzzles presented at the GameKnot site.

Chessgames.com (2018), on the other hand, presents puzzles as shown in Fig. 2 and has the following to say:

Figure 2: A chess puzzle displayed at the main page of Chessgames.com.

It's impossible to satisfy everybody with every puzzle, for what is too easy for one person is too difficult for another. To enable everybody to enjoy the puzzles, we arrange them so that their difficulty increases throughout the week. Monday and Tuesday puzzles are the easiest puzzles of the week; Saturday and Sunday are the most difficult.

Nearly everybody should be able to solve Monday and Tuesday puzzles, although beginners at chess might have to invest some time to see the solution. By Wednesday and Thursday even strong players are occasionally *stumped. Friday and Saturday puzzles can be notoriously difficult, and Sunday puzzles are often impossible to solve below the master level.*

Note that the concept of "difficulty" with chess puzzles is very subjective. Don't be surprised if some weeks you spend more time on a Tuesday puzzle than on a Friday puzzle--the escalating difficulty is just a rough guideline which cannot possibly apply to everybody. For most people, there are only two or three days a week when the puzzles are easy enough to solve, yet difficult enough to be challenging.

Suffice to say there is no official or standard method of classifying chess problems, even among major online chess communities. In section 2, we briefly review related work in the area. In section 3, we explain the methodology used. The article concludes with some experimental results and conclusions in section 4.

2 REVIEW

In the somewhat related psychology field, it has been known for over half a century that experts in chess likely have 'selective heuristics' that make it easier for them to solve chess problems rather than unusually large memories or 'processing speed' (Simon and Simon, 1962). The difficulty of (tactical) chess problems has also been considered with theories such as 'knowledge spaces' that relies on expert feedback and opinion (Albert et al., 1994). This, however, has little to do with automating the process of determining solvability (since human experts are required). More recent research (Bilalić et al., 2009) suggests further that experts in chess perform better when presented with problems within their area of specialization as opposed to those requiring general expertise; reinforcing and perhaps refining the 'selective heuristic' idea.

Generally, in psychology, it would appear the reliance on just one or more chess experts to determine the 'objective' difficulty of a chess problem persists even to the present year (Fuentes-García et al., 2019). We are of the opinion that this is quite possibly inaccurate and therefore inadvisable if used as a measure of how *human* players perceive such problems as a handful of experts hardly represent what the typical or even average chess player experiences when trying to solve a problem. More specific to the area of computer science or artificial intelligence (AI), Guid and Bratko (2013) proposed a measure of difficulty in chess that was based on 'sensible solutions' at increasing search depths using a chess engine. It was correlated well with errors made in world chess championship matches. However, in our view, the approach itself appears perhaps to depend a little too much on the choice of the engine (a critical factor overall?), the hardware it is running on and ultimately does not take into account the problem solving and perceptual ability of the vast majority of players and solvers, i.e. those who are *not* experts and more likely to benefit from trying to solve chess problems in the first place.

As far as we could tell, it was also not tested on actual compositions but rather positions taken from expert games. Stoiljkovikj et. al (2015) proposed the use of 'meaningful search trees' in order to estimate difficulty in chess problems but in our view, it too appears to tend to focus more on the capabilities of experts or experienced solvers. Another limitation or issue that stood out to us is the assumption or arbitrariness of what are considered 'reasonable moves'. This is not to say that these methods cannot *also* work to an extent (they clearly do); nevertheless ours is therefore provided as an alternative, if not something complementary or better. Other research in this area was scarce or deemed not relevant enough to the present topic. The interested reader may, however, refer to the broader reference sections of the aforementioned citations for additional material related to issues not explored here.

3 METHODOLOGY

In order to estimate the solvability (S) of a chess problem, we propose the following formula.

$$\mathbf{S = (EST \times 1000 + PC + SV + V) - ML, [if\ V > 100\ then\ V = 100 + V/100 - 1]}$$

EST = engine solving time (in seconds), PC = piece count, SV = Shannon value (of pieces),
V = (adjusted number of) variations, ML = move length

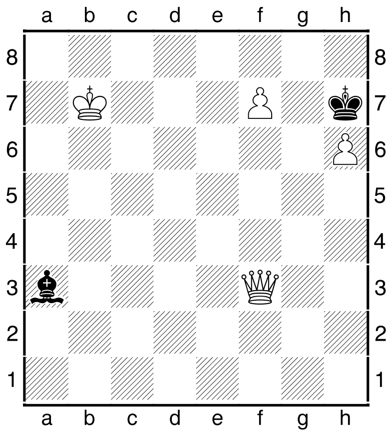
The Shannon (1950) values of the chess pieces (*Q:9, R:5, B/N:3, P:1*) is a commonly used measure of the value of the queen, rook, bishop, knight and pawn absent of positional considerations. In principle, any standardized measure of piece value could be applied. The 'variations' variable is significantly reduced if it exceeds 100 based on the assumption that anything more than this would probably be beyond the capability of even expert players to have considered in totality in any given position. Analogously, the 'engine solving time' variable is multiplied by 1000 to make it more comparable to the other variables given that chess engines running on even home computers today often only need microseconds to solve a puzzle. The raw (though objective) breadth and depth of the game tree (to a point) branching from the initial position was seen as undesirable as factors in the formula because most engines traverse them quite selectively depending on how they evaluate the initial or root position; the total number of nodes or positions in the branch are therefore effectively irrelevant with regard to solvability. The subtraction of the move length was to compensate slightly for the fact that longer mates are likely to be inherently more difficult for humans to solve. A higher solvability (S) value or score would therefore suggest the increased likelihood of a problem being perceived as more difficult by most players or the typical or average player.

We realize that alternatives to this formula and the proposed values or constants may also be just as viable (or perhaps even better) but we will leave that question to future work by researchers who may be interested in refining or improving upon the material we have presented here as a starting point. Fig. 3 shows a sample chess problem with its composition details, main line and solvability calculation to the left of the diagram, and its possible variations to the right.

Chesthetica v10.57
Selangor, Malaysia
White to Play and Mate in 3
2017.12.13 8:07:57 PM

1. Qf6 Bb2 2. f8=N+ Kg8 3. h7#

Solvability
= (0.078 x 1000 + 6 + 14 + 14) - 3
= 112



Qf6	Bb4	Qg7#		
	Bc5	Qg7#		
	Bd6	Qg7#		
	Be7	Qg7#		
	Bf8	Qg5	Be7	Qg7#
			Bd6	Qg7#
			Bc5	Qg7#
			Bb4	Qg7#
			Ba3	Qg7#
			Bg7	Qxg7#
			Bxh6	Qg8#
			Kh8	Qg8#
	Bb2	f8=N+	Kg8	h7#
	Bc1	Qg7#		

Figure 3: A sample chess problem, its solvability evaluation (left) and variations (right).

The engine used in this particular case was ChestUCI v5.2 (Huber, 2018) running on an *Intel(R) Core(TM) i5-4570 CPU @ 3.20GHz* desktop computer with Windows 10 Enterprise (64 bit) and 16 GB of RAM installed. The solving time was 0.078 seconds. As we can see, there are 6 pieces on the board worth 14 Shannon pawn units (Q = 9, B = 3, P = 1). There are also only 14 forced mate variations (including mates in 2 if Black did not play optimally). Finally, the number of moves is 3 given that the stipulation is mate in 3 (against any defense). The final solvability score for this chess problem is therefore 112. If the number of mate variations had been say, 165, then the 'V' value would have been calculated as 100 + 165/100 - 1 = 100.65. We have even come across problems seemingly just as difficult where the number of mate variations is in excess of 7,000. This is why the adjustment proposed seemed reasonable and even necessary at some cutoff point. This cutoff point, i.e. at which more possible variations no longer make a difference in terms of perceived difficulty to human players (of say, a particular playing strength), might be of interest to psychologists in some of their experiments and a worthy matter of investigation in itself.

In order to test the proposed formula for viability in estimating the solvability of forced mates, we compared mate in 3 (#3, for brevity) compositions by human composers (selected based on no other criterion than being a published #3), randomly-selected #3 computer-generated compositions, randomly-selected #4 computer-generated compositions and randomly-selected #5 computer-generated compositions. Each group contained 145 compositions each. The ones by human composers were taken from the FIDE Album 2001-2003 (Avner, 2011) and is the only data set of its kind we happened to have access to. This is the reason 145 compositions were also used for the other groups. The computer-generated compositions were by the prototype computer program, Chesthetica (Iqbal et al., 2016). Real game sequences (such as forced mates from tournament games) were not used as all the additional pieces unnecessary to the mate would likely add too much noise to the evaluations for a proper comparison to be made. Compositions tend to be economical in terms of the pieces used. Due to potential copyright issues, the PGN (portable game notation) files for these are not made public but may be requested from us via e-mail for research purposes, if necessary. The chess engine used was ChestUCI v5.2

running on an *Intel(R) Core(TM) i5-6300HQ CPU @ 2.30GHz* notebook computer with Windows 10 Home (64 bit) and 12 GB of RAM installed.

We did not test the formula against ratings by human experts because of what had already been alluded to in section 2, i.e. that expert opinions (usually just a handful anyway) do not represent in any objective sense the perceptual and cognitive difficulty experienced by the vast majority of chess players. In other words, if a handful of experts, even on average, say that problem A is more difficult than problem B, this is very unlikely to be the experience or opinion of the average or typical chess player (for whom an automatic estimator of solvability would be most beneficial). On the other hand, asking a certain number of average players (e.g. Elo around 1500) to evaluate or rank the difficulty of chess problems and using the average scores would likely suffer from a lack of understanding or knowledge that would also undermine any correlations or comparisons made to computational methods that depend on (strong) chess engines to assess the same chess problems.

4 EXPERIMENTAL RESULTS & CONCLUSIONS

The groups of chess compositions mentioned in the previous section would be expected to illustrate a varying level of difficulty, on average. This was the hypothesis, and the results based on the proposed formula are as shown in Table 1, rounded to one decimal place.

Table 1: Average solvability scores for different groups of compositions.

#3 Computer-Generated	#4 Computer-Generated	#5 Computer-Generated	#3 Human
116.3	148.0	225.4	233.8

A single factor analysis of variance (ANOVA) test was performed across all the four groups comparing the means and the differences were found to be statistically significant: $F(3, 576) = 137.1$, $p < 0.0001$. The results suggest that difficulty increases as the forced mate sequences get longer, which is to be expected. More interestingly, however, is that #3 compositions by human composers are not only more difficult, on average, compared to #3 compositions by computer but also more difficult than longer mate compositions (i.e. #4 and #5) by computer. This too is to be expected or is at least reasonable given the many more conventions and constraints put on human composers for their compositions to be recognized and published. The proposed formula can therefore be used to classify forced mate chess problems based on their difficulty. While there are many methods of delineating between categories such as 'easy', 'moderate' and 'difficult', we suggest using the percentile. For instance, anything that is <=0.33 can be considered 'easy', between >0.33 and <=0.67 can be considered 'moderate', and >0.67 can be considered 'difficult'.

In order for these values to be consistent or comparable across the chess problems, it is also important that the same chess engine and computer (i.e. the same software and processing power) be applied. The *EST* variable (see section 3) depends on this. The implication is that a different engine and different hardware could always be used with the same formula but doing so will call for a fresh evaluation of all the chess problems; therefore always using 'the best' or most up-to-date engine is not a necessity or a critical factor. Additionally, note that as new problems are added, some will likely shift from one category to the next but this will become less frequent over time as the database gets larger and the ranges stabilize. Table 2 and Fig. 4 depict how these values tend to adjust over time. The chess engine used was ChestUCI v5.2 running on an *Intel(R) Celeron(R) N4000 CPU @ 1.10GHz* notebook computer with Windows 10 Home (64 bit) and 4 GB of RAM installed.

A collection of the most recent 1,000 computer-generated chess problems, in chronological order, composed by Chesthetica were used. They contained #3s (35.5%), #4s (31.3%), #5s (25.1%) and studies (8.1%). The studies were automatically skipped by the engine in this case because the formula does not apply to them. They were not removed entirely from the start in order to more closely resemble a real-world application of the formula where a variety of chess problems may be collected or added from time to time (not all requiring a solvability estimate). This collection of 1,000 problems was broken into ten incrementing parts, i.e. 1-100, 1-200, 1-300 and so forth until 1-1,000. This would illustrate how the 33rd and 67th percentile values changed as more chess problems were added. Given the variability of the *EST* value (i.e. for a more stable determination of solvability) each chess problem was evaluated five times and the lowest solvability value used as the final score. This step is optional and some may prefer using the average score instead.

Table 2: Solvability score for mate (percentile) based on the ranges of compiled chess problems.

	1-100	1-200	1-300	1-400	1-500	1-600	1-700	1-800	1-900	1-1000
#3 (33)	209.0	202.0	205.0	205.0	205.0	212.0	200.0	241.0	238.0	209.0
#3 (67)	228.0	223.0	221.0	224.0	228.0	234.0	218.0	267.0	260.0	229.0
#4 (33)	250.0	233.0	229.0	226.0	240.0	238.0	215.0	260.0	254.0	225.0
#4 (67)	294.4	277.0	282.0	276.0	286.3	281.0	265.0	319.0	307.0	273.0
#5 (33)	296.0	304.0	311.0	314.0	320.0	312.0	282.0	338.0	328.0	288.0
#5 (67)	460.0	444.0	426.0	432.2	424.0	419.1	379.6	445.0	429.8	386.0

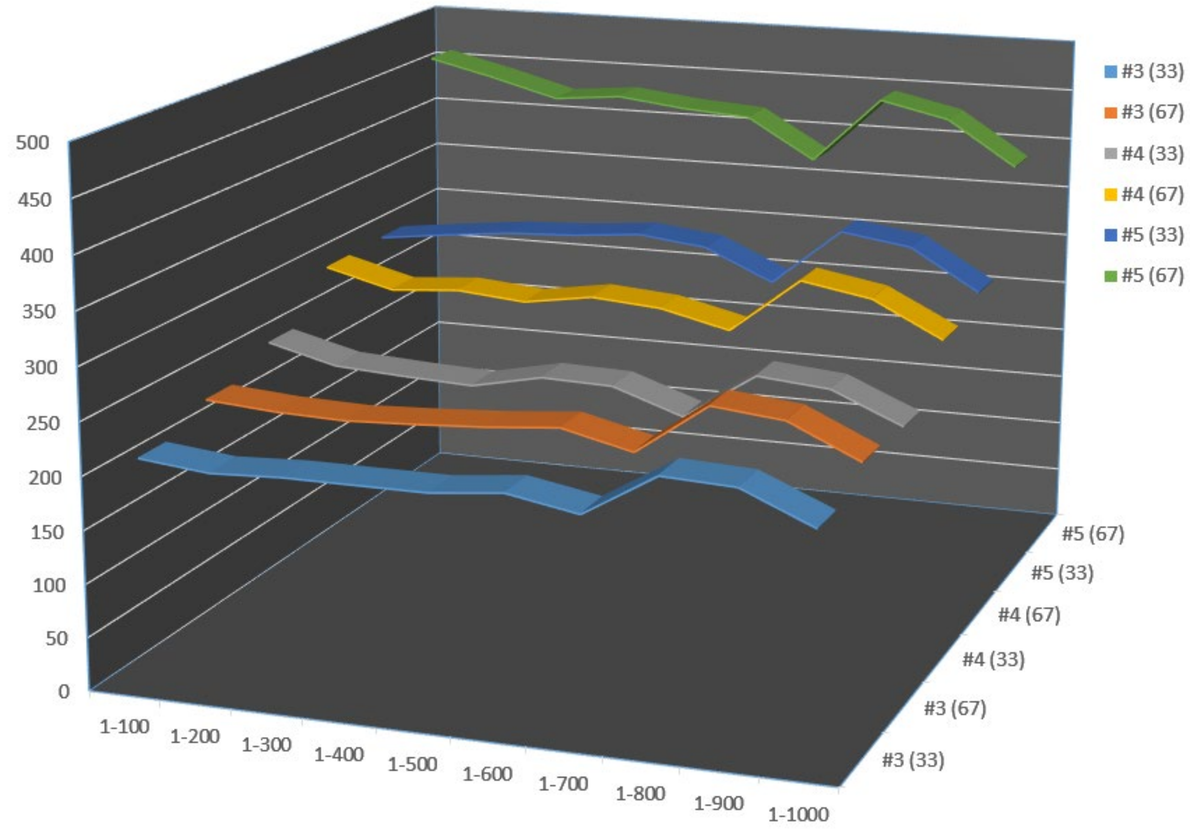


Figure 4: 3D visualization of the data in Table 2.

So, for example, we can see from Table 2 that after analyzing just 100 problems, the 33rd percentile for #3 had a solvability score of 209 whereas the 67th percentile had a solvability score of 228. This would mean that any new #3 problem evaluated could be classified as ‘easy’ if it scored <=209, ‘difficult’ if it was > 228 and ‘moderate’ for all other scores. By the time 1,000 problems had been added to the collection, the 33rd percentile for #3 was 209 and the 67th percentile was 229, showing very little change from the beginning (but more changes between, actually). The changes at the end are more significant for #4 and #5, however. Also, from Fig. 4, we can see a sudden rise or spike after adding problems 701-800 (i.e. between the 1-700 and 1-800 ranges). This can be expected to happen for whatever reason but then the values usually stabilize again.

With a sufficiently large collection of problems (e.g. 3,000 or more) it should be safe to assume that there will be very minor changes given additional (similar) problems such that any problem already evaluated in terms of solvability would not shift (e.g. from ‘moderate’ to ‘easy’) if evaluated again based on the updated solvability scores in percentiles 33 and 67. The updated amounts would not be much, if anything. Note that the solvability scores in Table 2 should not be compared with the values in Table 1 because those were generated using different hardware. Also, the newer computer-generated chess problems used here may be of higher (or lower) quality than the ones used in that particular experiment.

Ultimately, however, the perceived difficulty of a particular chess problem will depend on the solver, i.e. primarily their playing strength and experience (perhaps *in solving* chess puzzles or problems). The curious reader may peruse a growing collection of chess problems (Chess Intelligence, 2019) where the formula has been applied (using a consistent engine and hardware) to estimate solvability over the last few months. While studies (i.e. longer problems without a decisive result such as mate) may require a modification to the formula or different approach, we are fairly confident the formula as it stands should also work for forced mates longer than five moves, even though this was not explicitly tested here.

REFERENCES

Albert, D., Schrepp, M. and Held, T. (1994). *Construction of Knowledge Spaces for Problem Solving in Chess*. In Contributions to Mathematical Psychology, Psychometrics, and Methodology (pp. 123-135). Springer, New York, NY.

Avner, U. (2011). *FIDE Album 2001-2003*, Aachen, Germany. https://www.wfcc.ch/fide-albums. Accessed on 27th March 2019.

Bilalić, M., McLeod, P. and Gobet, F. (2009). *Specialization Effect and its Influence on Memory and Problem Solving in Expert Chess Players*. Cognitive Science, 33(6), pp. 1117-1143.

Chessgames.com (2018). *About Our Puzzles*. http://www.chessgames.com/aboutpuzzles.html. Accessed on 25th October 2018.

Chess Intelligence (2019). *Chess Intelligence (on Minds.com)*. https://www.minds.com/Chess_Explained. Accessed on 27th May 2019.

Fuentes-García, J. P., Villafaina, S., Collado-Mateo, D., de la Vega, R., Olivares, P. R. and Clemente-Suárez, V. J., 2019. *Differences Between High vs. Low Performance Chess Players in Heart Rate Variability During Chess Problems*. Frontiers in Psychology, 10: 409.

GameKnot (2018). *Chess Puzzles*. https://gameknot.com/chess-puzzles.pl. Accessed on 25th October 2018.

Guid, M., Bratko, I. (2013). *Search-Based Estimation of Problem Difficulty for Humans*. In: Lane H. C., Yacef K., Mostow J., Pavlik P. (eds) Artificial Intelligence in Education. AIED 2013. Lecture Notes in Computer Science, vol. 7926. Springer, Berlin, Heidelberg.

Huber, F. (2018). *Franz Huber's Program Projects*. https://fhub.jimdo.com. Accessed on 25th October 2018.

Iqbal, A. (2014). *Best 'Chess Constructs' by ChessBase Readers.* https://en.chessbase.com/post/best-chess-constructs-by-chessbase-readers. Accessed on 25th October 2018.

Iqbal, A., Guid, M., Colton, S., Krivec, J., Azman, S. and Haghighi, B. (2016). *The Digital Synaptic Neural Substrate: A New Approach to Computational Creativity*, SpringerBriefs in Cognitive Computation.

Shannon, C. (1950). *Programming a Computer for Playing Chess*. Philosophical Magazine. 41 (314).

Simon, H. A. and Simon, P. A. (1962). *Trial and Error Search in Solving Difficult Problems: Evidence from the Game of Chess*. Behavioral Science, 7(4), pp. 425-429.

Stoiljkovikj, S., Bratko, I. and Guid, M. (2015). *A Computational Model for Estimating the Difficulty of Chess Problems*. In Proceedings of the Third Annual Conference on Advances in Cognitive Systems ACS.

Wikipedia (2018a). *Chess Puzzle*. https://en.wikipedia.org/wiki/Chess_puzzle. Accessed on 25th October 2018.

Wikipedia (2018b). *Chess Problem*. https://en.wikipedia.org/wiki/Chess_problem. Accessed on 25th October 2018.