

Investigating the Role of Composition Conventions in Three-Move Mate Problems

Azlan Iqbal

College of Information Technology, Universiti Tenaga Nasional, Kampus Putrajaya,
Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
azlan@uniten.edu.my

Abstract. In improving the quality of their chess problems or compositions for tournaments and possibly publication in magazines, composers usually rely on ‘good practice’ rules which are known as ‘conventions’. These might include, *contain no unnecessary moves to illustrate a theme* and *avoid castling moves because it cannot be proved legal*. Often, conventions are thought to increase the perceived beauty or aesthetics of a problem. We used a computer program that incorporated a previously validated computational aesthetics model to analyze three sets of compositions and one set of comparable three-move sequences taken from actual games. Each of these varied in terms of their typical adherence to conventions. We found evidence that adherence to conventions, in principle, contributes to aesthetics in chess problems – as perceived by the majority of players and composers with sufficient domain knowledge – but only to a limited degree. Furthermore, it is likely that not all conventions contribute equally to beauty and some might even have an inverse effect. These findings suggest two main things. First, composers need not concern themselves too much with conventions if their intention is simply to make their compositions appear more beautiful to most solvers and observers. Second, should they decide to adhere to conventions, they should be highly selective of the ones that appeal to their target audience, i.e. those with esoteric knowledge of the domain or ‘outsiders’ who likely understand beauty in chess as something quite different.

Keywords: Chess, problem, convention, beauty, composer.

1 Introduction

A chess problem or composition presents potential solvers with a stipulation. For example, *White to play and mate in 2 moves*. They are typically considered works of art and therefore often described as ‘beautiful’ [1-5]. Composers usually adhere to many ‘composition conventions’ when competing in tournaments (or ‘tourneys’) in order to improve the quality of their problems. Examples include, *contain no unnecessary moves to illustrate a theme*, *avoid castling moves because it cannot be proved legal*, *possess more moves in the solution that are also of the ‘quiet’ type* and *create a deceptive setting for the solver*. A longer list is available in section 3.3.1 of [6].

Conventions are also useful as a kind of standard so that "*like is compared with like*" [7]. A case has been made for how not all conventions are actually prerequisites for beauty; see section 3.2 of [6]. Regardless, many composers and players tend to conflate or confuse convention with aesthetics such that a 'good' composition (one that adheres to conventions) is a more *beautiful* one. Award-winning compositions are therefore considered the most *beautiful*. We put this assumption to the test as it tends to lead to confusion in the world of composition and how others perceive their works. A selection of relevant material relating to computational aesthetics in chess can be found in chapter 2 of [6]. In the next section, we present our approach or methodology for this research. Section 3 contains details about the experiment and results. In section 4 we discuss these results. Finally we conclude the article in section 5 by presenting some closing statements and directions for further work.

2 Approach

In this research, we investigated three-move mate problems which are essentially positions where White is to play and can force mate in three moves against any defense by Black. We applied an experimentally-validated computational aesthetics model [8] to evaluate their beauty. The model has been demonstrated to be able to evaluate and rank this quality in chess problems in a way that correlates positively and well with domain-competent human assessment (not necessarily experts in the domain). It uses mathematical formulas as representations for well-known aesthetic principles and themes in chess. It also incorporates stochastic technology, i.e. the inclusion of some randomness. This means that the next time it evaluates the same composition the score could be *slightly* different. This is why for our experimental purposes, each composition was evaluated three times and the average score used. Additional information can be found in [8].

The model assesses primarily 'visual appeal' (ibid.) which is what the majority of chess players and composers with sufficient domain competence (e.g. the club-player level or higher) understand by 'beauty' in the game. This includes, for instance, tactical maneuvers like sacrifices or combinations with a clear objective such as delivering checkmate. These are relatively straightforward and easy to understand. 'Depth appeal', on the other hand, is associated more with strategic or long-term maneuvers that require more specialized or esoteric knowledge of the game, particularly in relation to chess problems. It is an open philosophical question which has more 'right' to the term 'beauty'.

A chess problem is expected to be legal (possible in a real game, however unlikely), sound (has a unique solution) and 'original' [9]. The first two are 'technical' issues that, if in doubt, can be determined quite reliably using retrograde analysis and a good mate solver engine. The third typically depends on the experience of the human judge. Computationally, it is virtually impossible to account for unless a database of all previous compositions is made available. Since originality can only be determined *ex post facto*, if at all, and has little to do with the 'inherent' beauty of the problem itself, its aesthetic consequences

are considered minimal given our experimental purposes. In addition to these three factors, ‘economy’ and ‘aesthetics’ are emphasized in the creation of chess problems (ibid.). The aesthetics model used takes into account all of them to varying degrees except for originality. There are conventions that relate to each of these five factors. In section 3.3.1 of [6], for instance, there are over 20 conventions listed. They therefore cover the spectrum of what makes a problem a ‘good’ one and it should be clearer now why simple adherence to them is often confused with what makes a problem ‘beautiful’ to human observers.

Chess problems composed by human composers for tournament or publication purposes can therefore safely be assumed to abide by the most number of conventions. Those composed by computer can be controlled to follow fewer or even no particular conventions at all. Forced three-move sequences taken from actual games – which technically ‘resemble’ chess problems – can be assumed to abide by few, if any, conventions since the players have no such concerns during the game. This provides a reasonable basis of comparison for the role of conventions when it comes to beauty in the game.

CHESTHETICA, a computer program which incorporates the model, was used to compose, as required, three-move mate problems [10] and also evaluate the aesthetics of such sequences. Computer-generated compositions tend to feature just one forced line and fewer, more easily identifiable conventions. The composing module of the program is separate from the aesthetics-evaluating one. One might suppose that the latter should be usable to aid the former but doing so has proven to be exceedingly difficult. The ability to rank beautiful pieces of art does not easily translate into the ability to *create* beautiful pieces of art.

3 Experiment and Results

For our experiment, we tested four sets of chess positions and their move sequences, i.e. three sets of compositions and one set of forced three-movers taken from actual games. The first set consisted of 145 compositions by human composers taken from the ‘FIDE Album 2001-2003’ [11]. This is all that was available for our purposes. The second set contained 145 compositions generated by CHESTHETICA using both a ‘random’ and ‘experience’ approach. The latter approach relies on a database of human compositions to determine piece-placement probability. Further details are available in [10].

These computer-generated compositions were constrained into adhering (randomly) to two, three or four conventions from a list of five namely, *no ‘cooked’ problems, no dual in the solution, no ‘check’ in the key move, no captures in the key move and no key move that restricts the enemy king’s movement*. A chess problem is ‘cooked’ when there is a second ‘key move’ (i.e. first move) not intended by the composer. A ‘dual’ occurs when White has more than one valid continuation after the key move. These conventions were all treated as equal and are among those more reliably detectable using a computer. Manual detection would have been too tedious, inconsistent and prone to error.

The third set consisted of 145 forced three-move sequences taken from real games between two chess engines (Rybka 3 vs. Fritz 8) under 1-minute time controls, known as blitz chess. The first engine, always having the white pieces, is the stronger of the two and they were configured to play until checkmate took place. These are therefore sequences that resemble composed chess problems but lack their typical qualities. The fourth and final set is the same as the second set except that the composing approach was entirely random (no ‘experience’) and there were no convention filters, meaning that the compositions generated need not conform to any of the five listed earlier. The number of compositions evaluated in all four sets was kept consistent to minimize potential statistical issues due to different sample sizes.

The compositions in the first set can safely be assumed to adhere to the most number of conventions because they were created for ‘official’ purposes such as competition and publication. It does not matter if the conventions they adhere to are the same as those generated by the computer because there is no known hierarchy of importance with regard to convention. The compositions in the second set adhered to anywhere between at least two and at least four conventions. This is because there may have been conventions other than the five listed earlier that were not tested for but happened to occur by chance, however unlikely. In any case, this second set can be assumed to contain fewer conventions than the first set. The third set in all likelihood contained the fewest conventions because there is no motive to compose anything of beauty in a real game, especially in a blitz match between two chess engines. The fourth set contained logically fewer conventions than the second and lacked ‘human’ composing knowledge. Table 1 shows the results.

Table 1. Mean aesthetics scores of the different sets of compositions and sequences

	Set 1	Set 2	Set 3	Set 4
Type	Human	Computer (2-4 conv.)	Real Games	Computer (0 conv.)
Typical Adherence to Conventions	High	Moderate	Low	Low
Mean Aesthetics Score	2.265	2.190	1.962	2.209
Standard Deviation	0.48	0.46	0.40	0.43

In comparing just the first three sets, the mean aesthetics scores were all different to a statistically significant degree using a single-factor ANOVA (analysis of variance) test; $F(2, 432) = 17.9$, $P < 0.001$. Given the reducing mean aesthetics scores across sets 1 through 3, this suggested that adherence to conventions did contribute positively to the perceived aesthetics of compositions or sequences in the game. However, using a two-sample t-test assuming equal variances to compare sets 1 and 2, and a two-sample t-test assuming unequal variances to compare sets 2 and 3, only the latter showed a statistically significant difference;

$t(282) = 4.47, P < 0.001$. This suggested that there was a limit to the role of convention when it came to beauty.

Set 4 (in contrast to set 2) came into play to provide a more realistic basis of comparison by using similar computer-generated compositions and ‘machine-adherence’ to convention. We found no statistically significant difference between the means of sets 2 and 4. This seemed to contradict the idea that conventions contributed positively to aesthetics. In the next section, we discuss these results in more detail.

4 Discussion

If we can accept that human compositions intended for competition or publication (set 1) adhere most to conventions, then it comes as no surprise that, on average, they are aesthetically superior assuming conventions do indeed improve the aesthetics of a problem. Using the same logic, computer-generated compositions that adhere to fewer conventions (set 2) should average lower, aesthetically. Even lower would therefore be sequences taken from real games (between chess engines, to boot) where conventions do not come into play at all (set 3). In comparing three such sets together, we find statistical evidence that this is all true.

However, in comparing them in pairs, we find that applying some conventions help¹. This is because the difference between sets 1 and 2 does not appear to be statistically significant yet the difference between sets 2 and 3 is. The increment in the number of typical conventions applied going from none (set 3) to some (set 2) is apparently significant aesthetically but from some (set 2) to many (set 1) is apparently not. On the other hand, if we contrast set 4 (also none) against set 2 (some), we find that the use of some conventions is also irrelevant. Sets 2 and 4 have more in common than any of the first three sets do with each other so the result here should not be ignored.

In order to reconcile the last finding with the rest, we need to assume that there is, in fact, a hierarchy to the many different composition conventions, especially with regard to aesthetics. In other words, some matter and some do not. Furthermore, among those that do matter, some are likely more important than others. This is why the five conventions that applied to sets 2 and 4 had little aesthetic effect. Yet, this little effect was sufficient in contrasting the aesthetic difference between sets 2 and 3 even though in the case of the latter, the relatively crowded board and lack of planned economy in the real game sequences probably worked against whatever aesthetic content they had that might have compensated for those five conventions.

Set 1, having the most in terms of convention, suffered an inverse effect. This means that there were probably some conventions that had little to do with aesthetics and others that had more; these cancelled each other out aesthetically which is why there was no difference in comparing sets 1 and 2. While we acknowledge that there may be other explanations or ways of interpreting the

¹ By ‘some’ we mean both ‘a few’ and ‘the right ones’ [12].

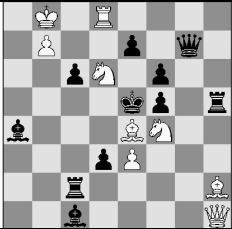
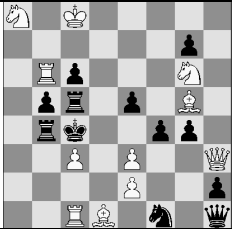
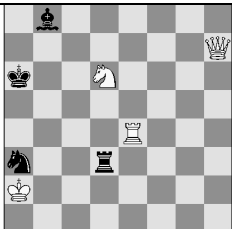
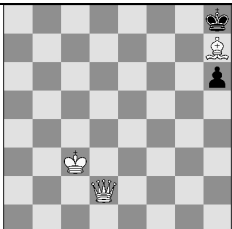
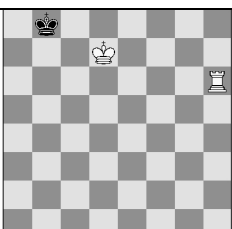
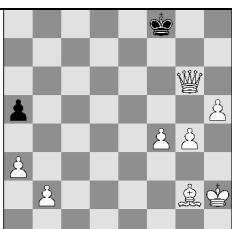
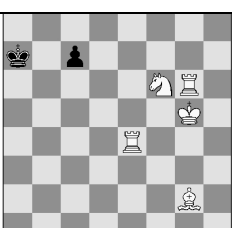
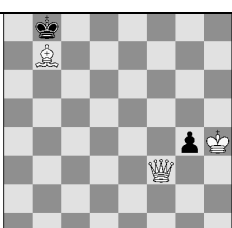
	
1. Ka8 Rg2 2. Nxd3+ Ke6 3. Nc5# (3.623)	1. Qh7 Kd5 2. Ne7+ Ke6 3. Qg6# (1.301)
Set 1	
	
1. Ra4+ Kb6 2. Qb7+ Kc5 3. Ne4# (3.894)	1. Qd7 h5 2. Qf7 h4 3. Qg8# (1.229)
Set 2	
	
1. Kc6 Ka7 2. Rh8 Ka6 3. Ra8# (3.100)	1. Qf6+ Kg8 2. h6 a4 3. Qg7# (1.038)
Set 3	
	
Kb5 3. Bc6# (3.565)	1. Qc6 g3 2. Bc8 g2 3. Qb7# (1.285)
Set 4	

Fig. 1. The highest (left) and lowest (right) scoring compositions/sequences from each set

results, the explanation provided above is what we believe is most likely the case and consistent with all the findings in this research. For the interested reader, Fig. 1 shows the aesthetically highest-scoring and lowest-scoring three-movers from all the four sets. Only the main lines and scores (in brackets) are given here.

Readers should not read too deeply into the precision of the scores because they are used primarily as ordinal data, i.e. for ranking purposes within a particular database of compositions or sequences. However, the aesthetics contrast between the highest and lowest scoring compositions and sequences should be evident to those with sufficient knowledge of the game.

5 Conclusion

Composition conventions are important in chess problems because they generally help composers create better works of art. However, the experimental work presented in this article suggests that not all conventions should be treated equally, especially when it comes to aesthetics or beauty as perceived by the larger community of domain-competent chess players and composers. These are people who need not be experts or possess esoteric knowledge of problems. Some conventions might even work against the typical conception of aesthetics in the game. The establishment of a standardized and recognized classification and ‘hierarchy’ of conventions with regard to their specific roles in chess problems would therefore be both practical and helpful. This would be useful to composers in tailoring their artworks to better suit their intended audience.

It is perhaps pertinent for experienced composers to realize that the term ‘beauty’, which they often use, actually carries different connotations outside the world of composition. This would help minimize confusion and frustration on both sides when trying to communicate chess problems as works of art. Any casual player – even with sufficient knowledge to recognize beauty in the game – who has bothered to go through the detailed analysis of an award-winning chess problem, would be able to relate to this. Just as any experienced composer who has had to explain clearly why such a problem is ‘beautiful’ would as well.

One direction for further work in this area would be a replication of the experiment but using instead endgame studies which are compositions of a different class. This would be to see if the findings are similar. Another direction worth pursuing would be the development of a formalized, objective method of classifying and ranking conventions with regard to their roles in compositions. The knowledge and possibly technology gained from all this would likely improve the quality of automatic chess problem composition [10] and add significantly to the wealth of artworks available to us [13].

Acknowledgement. This research is sponsored in part by the Ministry of Science, Technology and Innovation (MOSTI) in Malaysia under their eScienceFund research grant (01-02-03-SF0240).

References

1. Osborne, H.: Notes on the Aesthetics of Chess and the Concept of Intellectual Beauty. *British Journal of Aesthetics* 4, 160–163 (1964)
2. Humble, P.N.: Chess as an Art Form. *British Journal of Aesthetics* 33, 59–66 (1993)
3. Troyer, J.G.: Truth and Beauty: The Aesthetics of Chess Problems. In: Haller (ed.) *Aesthetics*, Holder-Pichler-Tempsky, Vienna, pp. 126–130 (1983)
4. Walls, B.P.: Beautiful Mates: Applying Principles of Beauty to Computer Chess Heuristics. Dissertation.com, 1st edn. (1997)
5. Levitt, J., Friedgood, D.: *Secrets of Spectacular Chess*, 2nd edn. Everyman Chess, London (2008)
6. Iqbal, M.A.M.: A Discrete Computational Aesthetics Model for a Zero-sum Perfect Information Game. Ph.D. Thesis, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia (2008), http://metalab.uniten.edu.my/~azlan/Research/pdfs/phd_thesis_azlan.pdf
7. Albrecht, H.: How Should the Role of a (Chess) Tourney Judge Be Interpreted? *The Problemist*, 217–218 (July 1993); Originally published as *Über Die Auffassung Des Richteramtes in Problemturnieren*, *Problem*, 107–109 (January 1959)
8. Iqbal, A., van der Heijden, H., Guid, M., Makhmali, A.: Evaluating the Aesthetics of Endgame Studies: A Computational Model of Human Aesthetic Perception. *IEEE Transactions on Computational Intelligence and AI in Games: Special Issue on Computational Aesthetics in Games* 4(3), 178–191 (2012)
9. Velimirovic, M., Valtonen, K.: *Encyclopedia of Chess Problems*. Chess Informant, Serbia (2012) ISBN 978-86-7297-064-7
10. Iqbal, A.: Increasing Efficiency and Quality in the Automatic Composition of Three-move Mate Problems. In: Anacleto, J.C., Fels, S., Graham, N., Kapralos, B., Saif El-Nasr, M., Stanley, K. (eds.) *ICEC 2011. LNCS*, vol. 6972, pp. 186–197. Springer, Heidelberg (2011)
11. Fougiaxis, H., Harkola, H.: World Federation for Chess Composition. FIDE Albums (June 2013), <http://www.saunalahti.fi/~stniekat/pccc/fa.htm>
12. Iqbal, A., Rashid, A.: Intended Multiple Interpretations. In: *Proceedings of the 4th Malaysian International Conference on Academic Strategies in English Language Teaching*, Shah Alam, Malaysia, November 27–28, pp. 97–101 (2012), http://metalab.uniten.edu.my/%7Eazlan/Research/pdfs/imi_azlan_aishah.pdf
13. Myers, D.: What Chess Games and Chess Problems Tell Us About Digital Games and Art. *Digital Creativity* 23(3-4), 260–271 (2012)