Computer Recognition of Aesthetics in a Zero-sum Perfect Information Game

AZLAN IQBAL¹ and MASHKURI YAACOB²
¹College of Information Technology, ²Vice-Chancellor’s Office
Tenaga Nasional University
Km 7, Jalan Kajang – Puchong, 43009 Kajang, Selangor
MALAYSIA

Abstract: - Zero-sum perfect information games are those where all the moves are known to every player. Examples include chess, go and noughts and crosses. This research intended to see if aesthetics within such a domain could be formalized for machine recognition since it is often appreciated and sought after by human players. For this purpose, Western or International chess was the most suitable because there is a strong body of literature on the subject, including its aesthetic aspect. Eight principles of aesthetics were identified and formalizations derived for each to form a cumulative model of aesthetics. A computer program that incorporated the model was developed for testing purposes. Two novel experiments were then performed comparing thousands of chess compositions (where aesthetics is generally more prominent) against regular games (where it is not) and the results suggest that computers can recognize beauty in the game. Possible applications of this research include more versatile chess database search engines, better automatic chess problem composers and computational aid to judges of composition and brilliancy tournaments. In addition, the methodology applied here can be used to gauge aesthetics in similarly complex games such as go and generally to develop better game heuristics.

Key-Words: - aesthetics, chess, game, intelligence, beauty, composition, zero-sum, heuristic, evaluation

1 Introduction

A zero-sum perfect information game is a useful research domain because in theory, the entire game tree can be computationally generated with nothing left to chance. This is easy in games like noughts and crosses but more difficult for chess or go because the number of possible positions (or nodes) is much higher. The interesting thing about some of these games is that there is more to them than just winning (often best accomplished through brute-force analysis). This means that concepts and ideas that are synergetic or beyond the simple rules of the game can be analyzed using a discrete or formalized approach. In chess for example, aesthetics is also important and appreciated not only by grandmasters but average players as well.

Garry Kasparov - arguably the world’s strongest player - is reported to have said, “I want to win, I want to beat everyone, but I want to do it in style!” [1]. Computers currently play chess at the grandmaster level and have even defeated the world champion but they cannot tell a beautiful move sequence or combination from a bland one. This is because the objective has always been simply to win [2-4]. Computers are hence unable to compose chess problems like humans do or simulate appreciation of the game. Chess literature adequately covers its aesthetic aspect (refer section 2) and the research presented here was intended to see if this information could be formalized for computational purposes, not unlike heuristics used in game engines.

The result is a model of aesthetics based on the proposed formalizations of beauty principles in chess. It is potentially capable of giving computers the ability to recognize aesthetics in the game comparable to the way humans do. Section 2 reviews some of the relevant contributions to the area. Section 3 details the proposed formalizations and aesthetics model. Section 4 presents some experimental results intended to validate the model. A discussion on the results and related issues appears in section 5. The paper concludes with a general summary of the ideas presented.

There are over 700 million chess players and composers worldwide [5]. Therefore the authors believe this research presents significant findings with respect to AI within at least the domain of chess itself. Extensions to other games or areas of research are not fully explored in the limited space available here but a brief discussion on it is presented in subsection 5.1. The information that
follows is therefore specific to chess given the inextricable nature of aesthetics to its domain.

2 Review

One of the earliest formal references to the aesthetics of chess was by former world champion Emanuel Lasker. In his book, “Lasker’s Manual of Chess” he devoted an entire chapter to it [6]. There he writes of the concept of “achievement” (e.g. winning material, space, the game itself) being important to aesthetics and that comprehension of the game, not necessarily mastery, is all that is required for its appreciation. Margulies, a psychologist, derived experimentally eight principles of beauty in the game from the judgement of experienced players, as follows [7].

1. successfully violate heuristics
2. use the weakest piece possible
3. use all of the piece’s power
4. give more aesthetic weight to the critical piece
5. use a giant piece in place of minor ones
6. employ chess themes
7. avoid bland stereotypy
8. neither strangeness nor difficulty produces beauty

Similar criteria have been mentioned in other sources with notable additions [8-15]. For example, paradoxical maneuvers (e.g. sacrificing pieces) and being economical in their use are considered aesthetically pleasing as well. Levitt and Friedgood include the concepts of geometry and flow as additional elements of beauty in the game [11].

Geometry implies graphic effects such as alphabets formed on the board whereas flow refers to forced play rather than many confusing alternate variations. Aesthetics is not limited to compositions and is also found in real games (e.g. tournaments), though less prominently [12-14]. Brilliance prizes are even awarded at certain tournaments to games that are aesthetically noteworthy either in full or part [15].

Even though most composition conventions (e.g. include variations, no duals, no symmetry) do not apply to real games, aesthetics is inherently shared between the two domains as long as the rules are the same [16]. Given say, direct-mate compositions (mate in $n$ moves against any defense), they essentially only differ with real games in terms of perceived beauty. Experienced players can often easily tell if a position looks like a composition because it is too “unusual” or “convenient” to have occurred in a real game. They are also generally in agreement with composers about what constitutes beauty in the game [17]. Computationally, aesthetics has been largely left to humans. For example, computers are capable of deriving forced checkmate combinations by constructing a complete database (e.g. from a set of desired pieces) and working backwards one ply (half-move) at a time but not capable of any “creative” activity per se [18][19]. It is left to humans to judge if the combinations are beautiful despite being conventionally acceptable or “correct” from a composition standpoint.

Composition conventions are often used to benchmark chess problems with little explicit emphasis on aesthetic factors [20][21]. The two (i.e. conventions and aesthetics) are distinct but not mutually exclusive [22]. Real games for example, also exhibit aesthetic properties but do not adhere to most composition conventions (usually in excess of 20 “rules”) [23]. Previously, mainly chess themes (e.g. Grimshaw, Pickaninny, direct battery), as a principle of beauty, had been weighted for the purpose of automatic problem composition and this was done by consulting one or two master composers [20][21]. The values ascribed to those themes - typically exotic ones used in chess compositions and seldom in real games - were arbitrary and based primarily on experience.

This meant that some themes were preferred over others and that some or all themes might have to be weighted again if new ones were added since their values were relative to one another. Additionally, all implementations of a particular theme were valued equally even though some configurations would no doubt be more beautiful than others [24]. Walls showed that beauty principles performed better than regular chess heuristics in solving certain types of chess problems [25]. He adapted a selection of Margulies’ principles as beauty heuristics but used them to guide a game playing engine instead of evaluating the principles themselves. They were therefore merely identified as either being present or absent in a particular line of play.

The formalizations used for those principles however, were rather rudimentary to minimize computational load. So in terms of say, distance (or using all of a piece’s power), a queen moving a certain number of squares across the board was considered just as beautiful as a rook or bishop. For this research, weighting individual principles through supervised or unsupervised learning was not suitable because reliable test data (i.e. aesthetically rated positions) is scarce. Varying implementations of a particular principle are then also difficult to properly account for [26]. In addition, it was
considered unnecessary since chess is a limited and precise domain with its own established measures and units that are not typically subject to personal taste in the way that say, images are. In the latter case, linear regression or classification can be used to individually weight aesthetic features since there are no agreed standards for rating them [27].

The approach taken in this research is more akin to how the aesthetics of music is sometimes calculated, where discrete representations (e.g. frequency of notes, intervals) of particular attributes (e.g. pitch, volume) are used to recognize beautiful compositions [28][29]. However, chess is a more limited and less culturally-dependent domain than music so formalizations based on established metrics are probably more reliable. The next section describes in detail the metrics, chosen principles and scope of analysis.

3 Methodology
In 1950, Claude Shannon explained how a computer could be programmed to play chess using estimated values (i.e. pawn units) of the chess pieces (K=200, Q=9, R=5, B/N=3, P=1) so that a score for every position in the game tree could be obtained based on the amount of material captured [4]. The king is actually of “infinite” value since its capture means losing the game but for practical programming purposes, it is often valued significantly higher than all the other pieces combined. Alan Turing is often credited along with Shannon for proposing a computational approach to playing chess. Two of his piece values however, were slightly different (i.e. Q=10, B=3.5) and are not as frequently used [30].

Computers could therefore be programmed to decide which moves were the most favourable from a material standpoint and play a reasonable game of chess based on the piece value property and pawn unit metric. Modern chess programs essentially still employ Shannon’s (Type B) method and rely on material as a primary factor for evaluation. To improve performance, piece values are sometimes changed by a program during the course of a game based on positional considerations [3]. For this research, the standard Shannon piece values were used; the king however, was valued at 10 pawn units to be aesthetically in line with the other pieces.

“Mating squares” or squares onto which occupation by an enemy piece would result in checkmate are also legitimate threats and valued equivalent to the king. In aesthetic analysis, winning is considered a prerequisite - no beauty is found in losing - and there is no intention of driving actual game play. Aesthetic evaluation of a chess combination is done ex post facto or in retrospect on the completed move sequence to determine how beautiful it was. The squares of the chessboard itself are used as a metric to evaluate properties like distance and piece power (mobility) because more powerful pieces tend to control more squares [31].

Distance is measured as the number of squares between two pieces on any line (i.e. ranks, files or diagonals). If there are three squares between two pieces, the distance is calculated as four. Piece power is interpreted here as the maximum number of squares a piece could possibly control on an empty board and was found to be: king (8), queen (27), rook (14), bishop (13), knight (8) and pawn (4). The pawn’s power is based on the fact that it can capture one square to the left or right and move forward one square or two for a total of four. Piece power is used to derive slightly different values for identical maneuvers performed by different pieces. It is based on their inherent and relative importance as generally perceived in the game.

3.1 Selected Principles and Scope
Based on the literature surveyed (refer section 2), eight aesthetic principles in chess were identified and selected namely: successfully violate heuristics, use the weakest piece possible, use all of the piece’s power, win with less material, sacrifice material, checkmate economically, spread out the pieces (sparsity) and use chess themes. Margulies’ 4th principle was not explicitly included because it simply means to emphasize the role of the “active” (e.g. checkmating) piece in a move sequence. His 5th principle used imaginary pieces not within the scope of Western chess while the 7th and 8th principles tend to rely on previous knowledge and experience so they could not be included.

Geometry was not included because it is very rare, even in compositions whereas flow is somewhat biased against compositions that typically feature many side variations (and are even lauded for it by convention). The goal of this research was to evaluate aesthetics in both domains (real games and compositions) but only in aspects that are equally applicable. Given the vast possibilities in chess and feasibility issues, aesthetic evaluation was limited to orthodox direct mate-in-3 move sequences. This scope of analysis permitted access to a wide selection of chess compositions and combinations from tournament games. The evaluation function of each principle was designed based on the metrics and properties (refer section 3) to score a theoretical maximum value of
approximately 1 so there would be no arbitrary preference given to any principle. There was nothing in the literature surveyed to suggest that some principles are inherently better than others. Only in extreme cases of certain principles is this limit markedly exceeded. Normalizing to 1 would involve multiplying the scores with an arbitrary value and was thus avoided. For brevity and by convention, white is always assumed to be the winning side.

Checkmates - though preferably forced (like in direct-mate compositions) - are also considered beautiful even if they are not forced. A beautiful mating combination in a real game for example, is often due to the oversight of the opponent. For that reason it might be perceived by humans as less beautiful but only upon deeper analysis and this would have little to do with the beauty of the moves themselves [15]. A composition of the direct-mate variety however, would in such a case be invalidated by convention for being “incorrect”, not unaesthetic. Selfmate problems for example, require that both sides cooperate to checkmate black, primarily because certain (aesthetic) effects are not possible with direct-mates [32]. The selected principles and rationale behind their proposed formalizations are explained in the following subsections.

3.1.1 Violate Heuristics Successfully
Heuristics in chess are typically general rules that govern good play. A move that violates one or more heuristics is considered paradoxical if it results in an achievement of some kind (e.g. checkmate). Given the scope, four heuristics were selected for evaluation: keep your king safe, capture enemy material, do not leave your own pieces en prise (i.e. in a position to be captured) and increase mobility of your pieces. Other heuristics such as control the center and avoid doubled pawns were not included because the effects of violation are not as obvious in the short-term [33][34]. A violation of keep your king safe was defined as moving the king to a square that makes it prone to check on the next move.

If the king’s destination is within the center four squares of the chessboard, it counts as a complete violation and scores 1 full point. The next surrounding 12, 20 and 28 squares are worth 0.75, 0.5 and 0.25 points respectively. This is because there is greater risk of exposure as the king approaches the center. Not capturing enemy pieces that are exposed and could be captured advantageously also counts as a violation. Given the depth and complexity of some exchange sequences in chess and the related positional dynamics, only undefended pieces or defended ones worth more than the capturing piece qualify. A non-capturing move or one that prefers a piece other than the most valuable available violates this heuristic. Pawns do not count as pieces worth capturing because they are usually not worth the diversion and fall short of what is required for a decisive advantage in chess (i.e. 1.5 pawns between grandmasters) [35]. The score for this violation is calculated as the sum of the value of uncaptured enemy pieces divided by the value of the queen. Therefore a full point is scored in cases where a queen or pieces of equivalent value are not captured in favour of some other move.

Like the previous violation, leaving your own pieces en prise applies only to pieces and not pawns. There is no violation if the move played captures an enemy piece worth more than the one left en prise or if the friendly piece is favourably defended (no potential loss of material). Exchanges were analyzed to a depth of two plies. The score is calculated as the sum of the value of en prise pieces divided by the value of the queen. The last violation is decreasing your own piece mobility. Usually, players try to control more squares with their pieces but sometimes the opposite is done and this can be paradoxical. For example, a queen or bishop may be moved to the very corner of the board behind some friendly pieces where its mobility is greatly reduced or moved to block several other pieces, reducing overall mobility.

The score is calculated as: \(\frac{w_1-w_2}{w_1}\). Here, \(w_1\) and \(w_2\) denote the number of legal moves for white in the initial and subsequent position respectively (assuming for a moment, black skipped his turn). Violation occurs if the score is positive. This principle is determined only after white’s first or key move because in compositions the first move usually points to the solution and by convention, the most surprising to solvers. The overall score for this aesthetic principle \(P_1\) is formalized as shown in equation 1. More heuristic violations \(n\) can be included as long as their maximum scores are in line with the others (i.e. \(\approx 1.0\)).

\[
P_1 = \frac{\sum_{i=1}^{n} v(h_i)}{n}
\]

\(v(h_i) = \text{value of a particular heuristic violation}\)

(1)

It is possible that the presence of more violations in a move would sometimes actually lower its overall aesthetic score for this principle since the average is used. This was necessary to keep the score consistent with the other aesthetic principles. In addition, it is arguably the nature of individual
violations that influences aesthetic perception of this principle more than the number of violations.

3.1.2 Use the Weakest Piece Possible
This principle means using the weakest piece possible to achieve a particular objective. Given the scope, it was refined to mean using the weakest piece possible to checkmate and therefore applies to the last move in the combination. The score is inversely proportional to the piece power of the checkmating piece. The formalization is given as:

\[
P_2 = \frac{4}{r(p)}
\]

\(r(p) = \text{piece power}
\]

(2)

The numerator is set to 4 so that if a pawn is used to checkmate, the score reaches its maximum of 1. In the case of a double checkmate (i.e. two pieces mating the king simultaneously), only the piece that moved counts. In the rare case of a two-way discovered checkmate, the weaker of the two pieces is chosen (refer Appendix).

3.1.3 Use all of the Piece’s Power
Using all of the piece’s power relates to its efficiency and can be interpreted as the number of squares a piece traverses in a single move. Traveling a greater distance is considered more beautiful than a shorter one. If a weaker piece (e.g. bishop) travels a certain distance, less of its power is wasted than if it were a more powerful one (e.g. queen). The bishop move is therefore considered more beautiful than the queen move. This principle applies to all the moves except those of the opponent because they usually work against the desired achievement (and hence aesthetics) of the winning side. The score is calculated as follows.

\[
P_3 = \sum_n \frac{d(p_n)}{r(p_n)}
\]

\(d(p_n) = \text{distance traveled by a piece}, \quad r(p_n) = \text{that piece’s power}, \quad n = \text{number of evaluation stages (i.e. each move by white + checkmate)}
\]

(3)

The knight, given its unique movement, defaults to a fixed 3 squares. In a mating combination, the distance between the checkmating piece and the enemy king in the final position is also evaluated in terms of this principle. It is possible in certain positions for the total score to exceed 1 (e.g. two maximal pawn moves + one knight move + mate using knight = 1.75) or fall significantly below it (e.g. two single square queen moves + one single square rook move + mate using rook right next to the king = 0.22). Like the previous principle, it applies to all combinations regardless of how beautiful or bland they might be.

3.1.4 Win with Less Material
This principle is considered aesthetic mainly because it is paradoxical. Usually, the side with more material is likelier to win. It applies only if black’s total material value exceeds white’s. The value is calculated as:

\[
P_4 = \left( \frac{b_n - w_1}{m} \right)_{b_n > w_1}
\]

\(w_i/b_j = \text{initial material of white/black}, \quad m = 38
\]

(4)

The denominator is set to 38 because this is the maximum amount of expendable material for an army (at least one pawn must be left) where checkmate is still possible, however unlikely. In such a case, black would have some material on the board that would blockade his own king and facilitate the checkmate. The possibility of such a combination is interesting in itself. The score is calculated in the initial position.

3.1.5 Sacrifice Material
Sacrificing material is also paradoxical. It is not quite the same as violating the heuristic of leaving your own pieces en prise. Rather, it applies more to exchanging your pieces in a seemingly unfavourable way in order to secure a decisive advantage or force a win. The “romantic” players of the late 18th and early 19th centuries often used bold sacrifices that were not always sound to impress spectators [36]. Former world chess champion Mikhail Tal, who considered chess first and foremost an art, was also known for intuitive sacrifices that gave rise to complications on the board and confused his opponents [37]. Since the late 20th century however, sacrifices are not as common in tournaments because players – having trained with computers - can usually spot weaknesses in them. Computer analysis can also quickly reveal similar flaws in compositions. Even so, sacrifices are still employed - even required in some positions - but are more calculated and
scrutinized than ever before. The “dramatic effect” of a sacrifice usually correlates with the amount of material lost so the function below (equation 5) is used to calculate the value for this principle. The material constant depends on the number of moves there are in the combination.

\[ P_5 = \frac{(w_1 - w_2) - (b_1 - b_2)}{m}, m \in \{9, 14, 19\ldots\} \]

\[ w_1/w_2 = \text{initial/final material of white}, \quad b_1/b_2 = \text{initial/final material of black}, \quad m = \text{material constant} \]

For example, a mate-in-2 sequence would have a material constant of just 9 because this (a queen’s value) is the most amount of material that could be lost to the opponent in one move. The checkmating move does not count because the opponent cannot respond. A mate-in-3 would have a constant of 14 since after the opponent’s second move, at most another rook (given the original piece set) could be lost and so forth. No sacrifices are possible for mate-in-1 positions and only positive values apply.

This function takes into account sacrifices of any number of pieces of any type, including adjustments for pawn promotions by both sides because the net difference in material at the end of the move sequence will reflect how much material was really lost. It would be misleading for example, to sacrifice a knight after the first move only to promote a pawn to a queen on the second. Negative values indicate that white actually gained material but this is not held against him because many mating combinations necessarily result in significant material loss by the opponent. They are however, less beautiful.

3.1.6 Checkmate Economically

Economy in chess ideally refers to using the minimum amount of resources – in most cases this means material - to achieve a particular objective. For the scope, the objective is to checkmate the opponent. This principle is therefore evaluated in the final position where economy is most often exemplified [38]. It is difficult to ascertain economy in the moves preceding the final position because they may contain sacrifices or in contrast, “quiet” maneuvers that are necessary but do not make much use of a piece’s power.

Economy can be formalized as shown in equation 6. The parameters are essentially based on the conventions used by the Bohemian “school” (i.e. style) of composition which is known for its emphasis on economy [39]. The other two schools are the Logical and Strategic [11][23]. These do not neglect economy but often make concessions in the interest of themes. Due to space limitations, a detailed explanation of this function and all its parameters is referred to in [40].

\[ P_b = \frac{\sum a_n f_n - \left( o + \sum s_n f_k \right)}{p} \]

\[ a_n = \text{control field of a particular active piece}, \quad f_n = \text{maximum control field of that active piece}, \quad o = \text{number of overlapping control field square}, \quad f_k = \text{standard king’s domain (i.e. 9)}, \quad s_n = \text{maximum control field of a particular passive piece}, \quad p = \text{number of friendly pieces on the board} \]

3.1.7 Spread out the Pieces (Sparsity)

Positions that are cluttered or crowded are generally considered less beautiful than those more spaced out [9]. Two important features when evaluating sparsity are therefore the number of pieces on the board and their proximities to each other. Even so, a position that requires more pieces should not necessarily suffer in terms of being sparse than say, an endgame position where pieces are inherently few. There are several ways that sparsity or its inverse, density can be evaluated (e.g. like pixels in a matrix, using quadrant density ratios, counting symmetries) but they do not translate as well to the chessboard [41].

For instance, a relatively “dense” quadrant of the chessboard may be considered sparse if there are only 3 or 4 pieces there because it is not practical or useful for them to be spaced out into different areas. There are also complications when we consider the centre 4x4 squares of the board as constituting a “fifth” quadrant because sometimes pieces are concentrated there. In fact, activity or checkmates at the center of the board are considered more beautiful than at the edge or corner [42]. Ideally, evaluation of this principle should be able to differentiate between positions to the point of sufficient sparsity, which is when a position is no longer cluttered or crowded enough to warrant improvement.

This is the point when the board configuration no longer plays a role in the evaluation. An example is say, just three pieces on the 64-square chessboard. One could imagine many different configurations of those pieces that would be considered sufficiently sparse. A more effective method to evaluate sparsity that works well with chess (and other similar board...
Surrounding pieces are those in the field of a particular piece (i.e. the squares immediately around it). Fewer pieces around a particular piece make the area appear sparser.

\[ P_7 = \frac{1}{\left( \sum_{i=1}^{n} s(p_n) \right) + 1} \]

\[ s(p_n) = \text{pieces surrounding a particular piece} \]

The field is used to ensure that if all the pieces have at least one square in every direction vacant, the position is considered sufficiently sparse. This field can be expanded to two squares or more for larger boards (e.g. 19x19 in go). The average number of surrounding pieces is used to provide a better general idea of how uncluttered a position is. One is added to the denominator to prevent a division by zero error where there are no surrounding pieces. Both black and white pieces are taken into account.

Figures 1a and 1b show the sparsity scores of positions taken from a composition and tournament game respectively. The higher the score, the sparser it is. Arbitrarily adding or removing pieces will not necessarily bring the score down or up. It depends on how they affect the board configuration. Evaluations of many different positions suggested that this function captured the perception of sparsity in chess better than alternative methods.

\[ A = \sum_{i=1}^{n} P_n \]

\[ A = \text{aesthetic value of a combination, } P = \text{aesthetic principle evaluation score} \]

The sum of scores for aesthetic principles present in a combination should in theory, be higher for beautiful ones. It stands to reason that attractive or “brilliant” move sequences in real games and compositions should contain not only a higher frequency of aesthetic principles but better instances or configurations of them which the formalizations can be analyzed at greater depths than possible with the full set of heuristics and subsequently aid move selection because sparse positions are generally associated with less complicated positions whereas dense ones suggest impending difficulties.

### 3.1.8 Use Chess Themes

Themes in chess are essentially good tactics. Common themes include the fork, pin and skewer whereas more exotic ones - used primarily in chess problems - include the Grimshaw, Pickaninny and Plachutta. The effective use of themes is fundamental to aesthetics in chess. Ten themes common to both compositions and real games were selected [11][43][44]. For the time being, only their presence was taken into account.

\[ P_8 = \sum T \]

\[ T = \text{occurrence of chess theme} \]

The score is calculated as simply the number of theme occurrences in the move sequence. The selected themes included the fork, pin, skewer, x-ray, discovered/double attack, zugzwang, smothered mate, crosscheck, promotion and switchback. A single move can contain more than one theme. Formalizations for each of these themes are under development.

### 3.2 Model of Aesthetics

The individual formalizations for the principles described above are insufficient for evaluating aesthetics in chess even given the scope, though they might be capable of identifying highlights of a particular move sequence. A model of aesthetics is therefore proposed in the form given below.
proposed are flexible enough to evaluate. The presence of more principles however, does not guarantee a high score (their individual evaluations may be low) and neither does fewer principles guarantee a low score (individual evaluations may be high). This model also leaves room for the inclusion of more aesthetic principle formalizations than the current eight.

4 Experimental Results
A computer program called CHESTHETICA was developed incorporating the aesthetics model for experimental purposes. The program does not possess any game playing intelligence but is capable of facilitating a complete match between two players. This was necessary to set the foundation for proper evaluation of all the aesthetic principles and detection of relevant themes. Two experiments were designed to see if the computer program would rate, on average, compositions higher than tournament games in terms of aesthetics; consistent with human perception of beauty in chess.

For this purpose, a random sample of 10,000 mate-in-3 chess compositions (mostly published and by professional composers) was compared against a random sample of 10,000 mate-in-3 combinations taken from tournament games (refer Appendix). Since aesthetics in chess tends to correlate with sound play, only games between master players (ELO rating \( \geq 2300 \)) were used. The ELO rating system is a widely employed method for estimating the relative skill of chess players where a minimum rating of 2300 usually qualifies for an FIDE Master title. Novice or intermediate play would inherently be less beautiful and bias the results.

While most master games tend to end with one player resigning (as opposed to being checkmated) the wide availability of games provided a sufficient resource for the kind required. Most resigned master games however, are not so close to checkmate. The tournament game combinations used in the experiments were not necessarily forced mates like in the compositions because this does not influence its aesthetic evaluation in any way. The important thing is that the mates were played out in full by humans and not generated artificially by a computer.

Figure 2 shows the results obtained. They have been sorted in descending order for clarity. The compositions scored an aesthetic mean of 3.04 (SD 1.05) whereas the tournament games scored a mean of 1.92 (SD 0.94). A two-sample t-test assuming unequal variances showed the difference in means to be statistically significant; \( t(19741) = 79.7, P < 0.001 \).

Since themes did not have individual formalizations, a second experiment was performed excluding them.

This time the compositions scored a mean of 2.34 (SD 0.72) compared to 1.41 (SD 0.55) for the tournament games. A two sample t-test assuming unequal variances showed that the results were also statistically significant; \( t(18776) = 103.5, P < 0.001 \). Including themes, the mean was 58% higher for compositions and without them, 66% higher. The implications of these results and possible applications of this research are discussed next.

5 Discussion
The statistically significant differences in means found between the overall aesthetic scores of chess compositions and tournament games suggest that aesthetics in chess can be recognized computationally. The wider disparity seen between the means of both groups without taking into account the presence of themes suggests that individual formalizations help capture aesthetics better. This does not mean that compositions necessarily score higher than real games in terms of beauty because there are always exceptions such as poorly composed problems and overrated combinations in real games.

Figure 2 clearly shows that there are combinations in real games that in fact score higher than some compositions. Nevertheless, a high score based on the aesthetics model proposed would likely point to a move sequence that humans would find beautiful. Experiments involving human players were not performed because their knowledge of what constitutes beauty in chess would be difficult to ascertain as reliable (e.g. like that found in chess literature). However, a positive correlation between scores based on the model and those of competent (not necessarily master) human players would
validate the model even further. This remains to be done. It is difficult to determine if shorter or longer move sequences would exhibit similar aesthetic scores because shorter ones tend to be quite simple (and limited thematically) whereas longer ones can be difficult for humans to follow. Comparisons between combinations of different lengths are not as reliable for the same reason [33].

Modifications or extensions to the model and individual formalizations could be applied where necessary to compensate for these possible discrepancies. Chess database search engines can incorporate the aesthetics model proposed to locate aesthetically pleasing combinations in vast databases of games for human appreciation and study. They can also employ the individual evaluation functions to gauge certain features in games. Automatic problem composers can use the formalizations presented to refine their fixed-value approach to aesthetics and to decide - without human intervention - which derived forced checkmates are the best.

In addition, chess composition and brilliancy prize judges might find some impartial assistance through this model when deciding on a winner [45]. Finally, complex compositions could more quickly be solved - and in certain cases even solved at all - if game engines also employed heuristics based on aesthetics. Traditional heuristics used to prune the game tree in conjunction with brute-force searching sometimes cause paradoxical but necessary key moves to be missed [25].

6 Conclusion
In this paper, formalizations for established aesthetic principles in chess were proposed and presented cumulatively as an aesthetics model for the game. These formalizations used mainly the inherent metrics and properties of the game instead of arbitrarily assigned weights [35]. The aesthetics model was incorporated into a computer program to compare large random samples of orthodox direct
mate-in-3 compositions and similar combinations from master-level tournament games. The results showed a statistically significant difference in their means suggesting that computers can use the model to recognize beauty in the game. The aesthetics model can be further enhanced by including formalizations of additional aesthetic principles and individual formalizations for a variety of chess themes. Work on the latter is currently in progress.

Chess literature places no emphasis on particular aesthetic principles so not weighting them individually minimizes bias. Applications of this research are most obvious within the domain of chess but extensions to other games of similar complexity are feasible. Researchers in other areas might also draw on the work presented here in creative ways since chess is a useful and popular domain of investigation. With sufficient processing power, it is quite possible that computers will one day be able to discover amazing and brilliant combinations for human aesthetic appreciation and study that would otherwise take centuries to occur in real games or be thought of by composers.

7 Acknowledgements
I would like to thank John McCarthy (Stanford), Michael Negnevitsky (University of Tasmania), Jaap van den Herik (University Maastricht), John Troyer (University of Connecticut), Malcolm McDowell (British Chess Problem Society), Brian Stephenson (Meson Database), GM Jonathan Levitt and IM David Friedgood for their comments and feedback. I would also like to thank the University of Malaya for their support and cooperation.

8 Appendix
Two-way discovered checkmate example.
FEN: 1r6/2p1R3/3k4/1P1P4/7r/8/6n1/1KBR4 w
(1. Ba3+ Rb4+ 2. Bxb4+ c5 3. dxc6++)

Chess problems obtained from Meson Database (74513 problems); http://www.bstephen.me.uk/chessproblems/meson/meson.html

FIDE Master tournament games obtained from ChessBase MegaDatabase 2007 (3512846 games); http://www.chessbase.com/shop/

References:


