A Taxonomy of Concepts for Evaluating Chess Strength: examples from two difficult categories

By Hans Berliner*, Danny Kopec**, and Ed Northam**

*Department of Computer Science, Carnegie Mellon University
**Department of Computer Science, University of Maine

Abstract
This paper is an attempt to classify that which is important in chess. Other authors have chosen to make classifications relating to, for example, the material on the board, or whether there exists an attack on the king. We find that this is sufficient for an introductory treatment of the subject of advantages. However, the chess specialist understands that two apparently similar positions may, in fact, be quite different since there may be a very important difference in the features of each. Earlier attempts to develop test suites for evaluating human and computer chess strength [Reinfeld, 1958; Hort & Jansa, 1980], although valuable, have had clear drawbacks in terms of their depth, range and number of positions examined. We present a taxonomy of positions in chess that require special knowledge. The taxonomy is what drives our selection of positions, and not vice-versa. It is easy to understand what a passed pawn is, and a bit of classic advice such as "Passed pawns must be pushed" makes use of the simple metric that a passed pawn becomes more valuable with each advance. However, there are outside passed pawns, protected passed pawns, blockaded passed pawns, and passed pawn masses. Each requires its own understanding, and frequently other features of a position can cause great variation in what may at first appear to be positions that should be treated very similarly.

Our taxonomy is per force preliminary, as a work of this kind is essentially never complete. However, this is a beginning of a classification of the essential elements and how their interaction causes
important concerns. We conclude with a set of annotated example problems from two specific categories that require the user of this paper to demonstrate understanding of critical concepts to achieve the maximum result. These examples are the beginning of the construction of a large and comprehensive test suite which will be comprised primarily of positions which are tactically deep or strategically hard.

Introduction

During the time since 1965, that serious efforts have been under way to develop strong computer chess programs, a great deal has been learned about the structure of chess. Such things include the fact that brute-force searchers are very tenacious defenders, and that much that humans make a great deal of can be found, essentially en passant by a searching program. The reason that ideas such as board control do not appear in major chess programs is that board control correlates very highly with piece placement. If one side manages to place its pieces well while the other does not, this almost certainly means that the former side dominates the board in such a way as to prevent the opponent from achieving normal good squares for his pieces.

For such reasons the leading programs since 1980 appear to be deficient in knowledge. We understand how depth of search is related to good tactical play. However, a program that does not understand the weakness of a doubled isolated pawn will probably have to search to depths up to 40 ply to discover this fact from more primitive features such as material. Thus, even prodigious searchers require some knowledge (Berliner, 1984). Humans, who search significantly less than computers, require more knowledge.

However, some knowledge is required by all those who do not plan to search to find mates and stalemates. What computer chess has done is to shed considerable light on what the subject of advantages is all about. Certain advantages such as board control are merely correlates of other advantages. Pawn structure advantages are not. Further, there seem to be advantages that depend very much on a great deal of context. This is the subject of study of the "graduate school of chess knowledge" which elevates a player from class "A" to International Class.

Despite the ability of the top few programs to defeat all but the best few hundred players in the U.S., there remains considerable skepticism among artificial intelligence researchers (see e.g. Donskoy & Schaeffer, 1989; Kopec, 1990) as to whether these results represent competence or performance driven success.
The matches between World Champion Gary Kasparov and Ex-World Champion Anatoly Karpov against DEEP THOUGHT (October 1989, 2-0, and February, 1990, 1-0, respectively, won by the humans) suggest that there is still some way to go before top programs can seriously challenge the World Champion.

Usually performance can be achieved in two different ways:
1. By essentially defeating all those rated below and losing to those rated above.
2. By achieving non-uniform results against a spectrum of players that averages to a given rating.

Usually, the more erratic a player's performance (class 2 above), the more "promising" he is, and the earlier he will make his upward move by acquiring the necessary remaining skills. Players in class 1 are thought of as unimaginative and able only to wield a tactical cudgel. Thus, the success of top computer chess programs on the Elo rating scale is much more in accord with the former notion than with the latter. However, programs, in defiance of the "pure tactician" mold, have shown the ability to find unique, sound, beautiful, and important ideas by "totally inhuman methods" with relatively little knowledge.

While there is substantial evidence to support the contention that top programs can perform (obtain results) at the strong master level (namely their Elo ratings), there is little evidence to suggest that top programs have a deep understanding of the profound strategical consequences of a given move -- in this sense programs lack competence, for they will still make moves in a particular position which can be clearly assessed as bad according to elementary or classical heuristics.

We do not wish to sanctify knowledge for its own sake. Knowledge, like anything else in a chess playing entity must be able to pay its own way, and only "useful" knowledge should be acquired. Thus, while researchers into human thinking may find very large amounts of knowledge there, we will attempt to show that computers have an excellent method for distinguishing between useful and superfluous knowledge.

**Background and Methods**

The thought processes of chess players were studied by de Groot (1965) and he determined that Grandmasters look at somewhere between 50 and 200 nodes in the process of deciding on the best move in a given position. Chase and Simon (1973) demonstrated that it is the "chunking" of familiar chess specific patterns (groups of pieces) which deems Grandmasters superior to novices in their ability to recall a position from short term memory. They,
and later Nievergelt (1977) estimated that the number of chess-specific concepts stored in a master's head is somewhere between 50,000 and 100,000. Further work by Kopec and Bratko (1982) and Berliner and Campbell (1984) gave evidence for the important role played by pawn structures and pawn groups in strong players' abilities to determine what is critical in a chess position. However to this date, we know of no serious effort to develop a comprehensive taxonomy of important chess ideas. It is clear that many important chess themes and ideas have been subsumed by the tactical ability of top programs.

World Champion Kasparov believes that as programs get stronger, so will humans always be able to find new ways to exploit their weaknesses. Considering that he has the highest chess rating ever achieved (2800+), it seems worth taking his point of view seriously. It is clear that in chess there is a hierarchy to the vast number of ideas, heuristics, concepts, and even first order principles such as "In the Opening it is important to develop your pieces." It is also clear that many decisions by domain specialist experts (grandmasters in chess) do not follow from first principles. Grandmasters, like experts in any domain, know when to break the rules. They know when pieces should be moved more than once in the Opening neglecting the development of other pieces. Another example is the classic advice "Passed pawns must be pushed." However, when does a passed pawn become weak after it is pushed? As general rules for the novice and student, such heuristics are excellent for learning and improving play. There are, nonetheless, further refinements of general heuristics which require special knowledge about, for example, outside passed pawns, protected passed pawns, blockaded passed pawns and passed pawn masses, amongst others. More abstract chess concepts such as king safety, weak square complexes, notions of defense, time, space, and connectivity become harder to quantify, and bring to the fore issues of meta-concepts -- that is, how should concepts and their interaction be quantified?

Earlier attempts at chess test suites (i.e. Reinfeld, 1958; Kopec & Bratko, 1982) have either been put together for a different purpose or were not wide-ranging enough to provide a comprehensive test of chess understanding. Results on the Bratko-Kopec Test showed a strong correlation with ratings, especially for humans. For computer programs, none of which had at the time achieved master ratings, results on the 12 tactical positions were higher than would have been predicted by their ratings, but performance on the 12 lever test positions clearly indicated that this was an area where more domain specific positional knowledge was needed, i.e. tactical depth could not compensate for the lack of this knowledge. Further evidence of the problems presented by lever positions was demonstrated by Marsland (1989) from the results of his administering the test to the applicants in the 6th World Computer Chess Championship in Edmonton. The criticism that 24 positions
cannot reasonably be a representative sample from the space of the estimated
$10^{43}$ chess positions is certainly valid.

The taxonomy of chess ideas being presented here is only the beginning of a
classification process which in some sense can never be complete. Our
ultimate goal is to identify as many relevant classes as realistically feasible
with examples across all phases of play. Problems in the Opening will be
based on "common knowledge" themes derived from the vast repertoire of
human experience. The choice of Middlegame positions is based on difficult,
deep or complex tactical themes as well as deep-rooted, critical or refined
strategic ideas. Endgame classifications are focused on demonstrating the
importance of domain specific knowledge and distinguishing between similar
looking, but different positions, or on illustrating the depth belied by simple-
looking positions.

Chess programs may be viewed as highly tuned expert systems. Although
results from tournament play are quite revealing, an extensive (but not
necessarily exhaustive) test suite, such as the one we are constructing, is
necessary to identify, in an organized manner, the important concepts that still
elude computer chess programs. This test suite, due to its method of design
by taxonomy, will not only test chess strength, but will also help to define
what specific challenges lie ahead before the World Chess Championship can
be legitimately challenged by a program. In their final state these test
positions will be available in machine readable form. The taxonomy of
positions is continuously evolving across the three phases of play, Opening,
Middlegame, and Ending, and its current state is presented in Figure 1 below,
with 10 sample positions to follow.

Our focus in this paper, for purposes of illustration, are two specific Opening
categories, O5: Multiple piece moves to gain advantage and O13: Qualitative
vs. Quantitative Development. Positions in these two categories are
particularly appropriate for this test set because they exemplify some higher
level knowledge of when to break the rules -- that is they do not follow from
first principles. For example, we are taught to maximize minor piece
development with every move in the Opening. That is, to distribute
development amongst forces as uniformly as possible. This heuristic serves
very well as a general rule, and will by and large produce sound chess play.
Development is usually a good way to evaluate who, if anyone, has the
advantage in the Opening of a chess game. Yet there are instances where
pieces are developed with no clear purpose, perspective, or future. Despite a
countable deficit in development, one side's pawns may deem the opponent's
pieces futile for any active play. This illustrates the distinction between
qualitative and quantitative development. That is, how many pieces one has
deployed may not be as important as where they placed and what their future
potential is.
The second major counterexample category which we highlight is "multiple piece moves to gain an advantage". Every chess novice is taught the major principle of correct play, that one should distribute development uniformly amongst one's forces. That is, quite simply, "Don't move a piece twice in the Opening before you've developed all your forces unless you must. Nonetheless there are many noteworthy instances where finding and playing such moves distinguishes incisive attacking play which poses problems for the opponent from routine bland play.

In the context of computer chess play, where the trend is for most programs to have vast Opening libraries, one wonders how successful programs, relatively unconstrained by the rules of sound play, would be in finding such exceptional, multiple piece moves which distinguish this category of positions.

To make the test suite meaningful, we have attempted to produce a uniform scoring system that evaluates both the judgement of the value of the position and the ability to find the correct way to proceed. The points achievable by the user will vary according to the difficulty of the position, and the scoring takes account of things such as alternate ways of proceeding. We conceive of these positions being used by both humans and machines. In both cases, the attempt is being made to identify some facet of chess understanding which may be lacking. Every position falls into one of five class values according to the outcome with best play.

A Taxonomy of Chess Ideas

OPENING CATEGORIES (O)

O1. Gambits Accepted or Declined
O2. Return of Material at correct time
O3. Pawn for Development (e.g. Poison Pawn Variation, Najdorf)
O4. Bishop Outside/Inside pawn chain
O5. Multiple piece moves to gain advantage
O6. Good and Bad (attackable) Centers
O7. When to (and not to) fianchetto
O8. The Options Principle
O9. Critical pawn moves for space.
O10. Bishops for Knights (good and bad examples)
O11. Choice of where to develop B or N.
O12. Avoiding "shutout" of a piece(s)
O13. Qualitative vs. Quantitative Development
O14. Loss of Time due to multiple piece or pawn moves
MIDDLEGAME CATEGORIES (M)

Pawn Play (MP)

MP1. Pawn advances with idea of opening files (involving pawn exchanges)
MP2. Minority Attack
MP3. Preserving Pawns on side chosen to win on (e.g. Outside P.P.; Potential Outside P.P.)
MP4. Pawns on d5 or e5; center pawns blockaded or not.
MP5. Structures related to minor pieces.
MP7. Advances in front of king.
MP8. Advancing for space.
MP9. Keeping or relinquishing tension.
MP10. Attacking with pawns.
MP11. Central pawn mass.
MP14. Pawn Structures:
   (a) Doubled Pawns
   (b) Meaningful Majorities (Handling)
   (c) Isolated
   (d) Backward
   (e) Passed
   (f) Hanging
   (g) Chains

Piece Play (Positional) (MPPP)

MPPP 1. Normal material values do not apply
MPPP 1a. Exchange Sacs
MPPP 1b. Queen Sacs
MPPP 2. Piece Sacrifices
MPPP 3. When to give up the bishop pair; which one, how.
MPPP 4. Superior Knight
MPPP 5. Position of Rooks
MPPP 6. Choice of Rook Placement (where options exist)
MPPP 7. Which side to castle on (if at all)
MPPP 8. Piece Regroupings (R,B,N,Q)
MPPP 9. N on rim/ good/bad
MPPP 10. Trapped Pieces
MPPP 11. Freeing Trapped Pieces
MPPP 12. Bishops of opposite colors; attacking with; exchanging queens with;
MPPP 13. King walk
MPPP 14. Removing (trading off) the key defender of an opponent's position.

Piece Play (Tactical) (MPPT)

MPPT 1. Desperado
MPPT 2. Double Attack
MPPT 3. Hanging Pieces
MPPT 4. Interference
MPPT 5. Decoy
MPPT 6. Overload
MPPT 7. Skewer
MPPT 8. Trapping
MPPT 9. Pin

ENDINGS (E)

E1: R + P Endings (RPE):
RPE 1. Rooks Threatening Pawns
RPE 2. Rooks behind, on side of, in front of P.P.
RPE 3. Connected P.P.
RPE 4. Position of defending king
RPE 5. Superior Pawn Structure

E2: B vs. N (superior/weak)
E2a: B vs. B Same color, bad Bishop.
E3: Complex Pawn Endings
E4: Q + P endings
E5: R + B vs. R + N
E6: R + P's vs. Two Minor Pieces
E7: R + B + B + extra pawn(s) vs. R + R + N
E8: Mating attacks in endings
E9: General: Not how many pawns, but their position counts.
E10: Exchange up with pawns on board.
E11: Elementary Endings
   (a) R + P vs. R
   (b) R vs. N
   (c) R vs. B
   (d) Q vs. R
   (e) Q vs. BB
   (f) BB vs. N
   etc.
Example Positions

CATEGORY O5 EXAMPLES

position O5/1

position O5/2

position O5/3

position O5/4

position O5/5

CATEGORY O13 EXAMPLES

position O13/1

position O13/2

position O13/3
Solutions to Positions

The five position class values are as follows: (1) White is winning (2) White is superior (3) The position is equal (4) Black is superior (5) Black is winning.

Category 05 Examples: Multiple Piece Moves to Gain an Advantage

Class value = 1. Difficulty points = 3 for move and class value.

2. o5/2 Black has just played 10. ...O-O. This is an error because it allows the strong retort 11.Ba3 which Black probably did not expect since the WB has just been developed on b2 on the previous move. Now the BB on e7 is undefended and the p/c5 is pinned. After the impending trade of dark-squared bishops White can look forward to the exploitation of Black's bad light-squared bishop. A piece has been moved again, just after being developed, in order to achieve a higher positional goal.
Class value 2. Difficulty points 1.

3. o5/3 White has committed a classic Opening error in allowing the strong 7. ...Ng4. If the B/e3 moves (i.e. to g5 or c1) then 8. ...Qb6 is very strong. Otherwise Black takes on e3, destroys White's p-structure, and follows with ...Ne5 and ...e6, etc.
Class value 4. Difficulty points 1.

4. o5/4 The start of Alekhine's original opening masterpiece against Wolfe in which the WQ moves 4 times in the first 11 moves to gain a great advantage. 7. Qa4+! If Black plays Nbd7 then 8. Ne6 is very strong; and
if 7. ...Bd7, as occurred in the game, then simply 8. Qb3 is strong. The cramping effect of the p/d5 is already felt. Class value 2. Difficulty points 1.

5. 05/5 The famous and remarkable Ulvestad Variation of the Two Knights Defense (5. ...b5) which gained a serious test in Estrin - Berliner, 5th World Correspondence Championship, 1969. Black intends to meet 6. Bxb5 with Qxd5 leading to great attacking chances on the a8-h1 diagonal. The idea of 6. Bf1! is to guard g2 and to meet 6. ...Qxd5 with 7.Nc3, and if 6. ...Nxd5 7. Bxb5 follows. Class value 3. Difficulty points 2.

Category 013 Examples: Qualitative vs. Quantitative Development

6. 013/1 This is a continuation from position 4 above, Alekhine-Wolfe, Pistyan, 1922. It is a unique example of when to break a principle. 1. d4 d5 2. Nf3 c5 3. c4 cxd4 4. cxd5 Nf6 5. Nxd4 a6?! 6. e4! Nxe4. Now 7.Qa4+! Bd7 8. Qb3 Nc5 8. Qe3! g6?! 9. Nf3 Qc7 and now 10. Qc3! The multiple Q and N moves have kept Black from completing his development, and now he must decide between a permanent weakness by f6 or renouncing castling by Rg8. Clearly, the rule, "Do not move a piece more than once in the opening" was one which Alekhine knew exactly when to break. The position also belongs in this category, 013, Quantitative vs. Qualitative Development, as Black has more pieces developed but is losing the Opening battle. Class value = 1. Difficulty points = 3 total for move and class value.

7. 013/2 Qualitatively, Black has the lead in development, but White can effectively contain the Black pieces with 9.f3. Instead, 9.g5? is an error, weakening squares and offering counterplay via 9. ...Nh5. If 9. ...h5 then 10.g5 is now strong as Black's pieces stay cramped. Class value 2. Difficulty points 2

8. 013/3 An excellent example from this category of positions (013), as despite White's having two developed pieces against Black's four, White has a great advantage. Black's pieces are cramped and can hardly move. 23.a4!, as played in Alekhine - Fine (AVRO, 1938) prevents Black from getting any respite via ...Na4 and ...Nc5. Note how White's pieces and pawn contain Black's. Class value 1. Difficulty points 2

9. 013/4 A position from the Four Knights Variation of the English Opening which arises after: 1.c4 e5 2.Nc3 Nc6 3.Nf3 Nf6 4.e4 Bc5? when Black gains in development but is rudely awakened with 5.Nxe5! leading to a fork trick and an advantage for White after lines like: 5. ...Nxe5 6.d4 Bb4 7.
dxe5 Nxe4 8.Qd4 Nxe3 9. bxc3 Ba5 (9. ...Be7 10.Qg4) 10.Ba3 etc. Although appealing, 4. ...Bc5 is a somewhat crude development. Class value 2. Difficulty points 4 for move and 2 for continuation.

10. 0-13/5 Another position which can easily arise from the English. Again, in pure development Black is ahead. However more subtle, positional Openings like the English are poignant in demonstrating the contrast between ordinary development and development with a purpose. Black's B/c5 is misplaced and a target while its counterpart on c1 simply waits for its indicated and correct development. In any case Black must now play ...a5 to take some of the sting out of b4. Class value 2. Difficulty points 2.

Acknowledgements

The authors especially wish to thank A. John Ostuni for technical assistance in preparing the manuscript and Robert Cicogna for assistance in its electronic transfer.

References


