WHAT CONSTITUTES OPTIMAL PLAY?

R. Seidel
Institut für Psychologie
Technische Universität Berlin

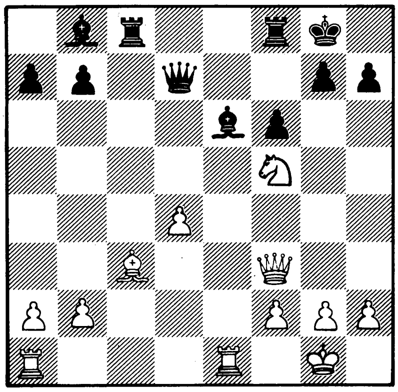
OPTIMALITY AND THE SHORTEST PATH

In the last years, it has become standard to define the shortest-path solutions as the "optimal" one, whereas other solutions are at best classified as "correct", as explicitely proposed by Bramer (1980). Recently, Roycroft (1985) raised some intuitive doubts on the adequacy of this convention in endgame theory and data-base construction. Since the issue of optimality indeed has an impact on the development of knowledge-based programming and chess theory in general, it seems to deserve a deeper investigation.

Normally, "optimal" - as a predicate of a solution - refers to some criterion which such a solution is expected to meet. For example, a computer program may be optimized to minimal running time or to minimal storage requirements or to maximal readability to a human, and in every case a different program may turn out to be optimal. What might be the reason for choosing the shortest-path solution in chess as <u>the</u> optimal one? Admittedly, there is some first-glance plausibility for this option (I myself had adhered to it for some time), but nevertheless it seems to be a mere <u>ad hoc</u> decision: the constructor of an endgame algorithm frequently is confronted with the existence of more than one winning move and has to decide which one to trace as the main variation; looking for some <u>objective</u> rule, the shortest-path criterion offers itself naturally.

The chess rules, the only obligatory benchmark of what has to be done in chess, provide no relief. Except for the 50-move rule (which should be abandoned anyway), a player is not at all obliged to pursue the fastest possible win nor will he be rewarded for it in any manner. (In passing: what about a new tournament rule such that a win will be scored higher proportionally to the shortness of the game?) Thus the chess rules do not justify considering a shortest-path solution as optimal.

Let us look at a practical play and consider what light it throws upon the question of optimality (See Diagram 1) (Plesse, 1985).



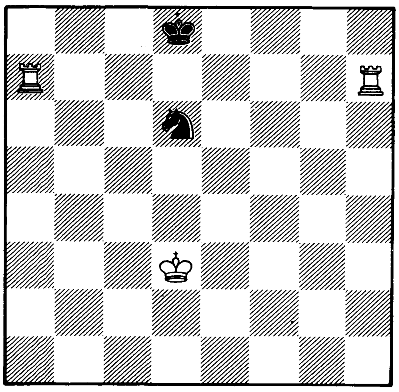
Chiburdanidze-Wiese
Tbilisi, 1984
White to move
DIAGRAM 1

The continuation in this position was: 1. Rxe6 Qxe6 2. Qg4 and Black resigned because she will lose the Queen by the double threat 3. Qxg7 mate or 3. Nh6+. The point of interest is Black's reply 1. ... Qxe6. Surely, no commentator would condemn this move as a blunder - there is no black move at all saving the game. But there are moves to avoid the immediate loss of the Queen (and of the game soon afterwards), though these moves would be at the cost of accepting that the Bishop is lost, after which White still has a strong attack. A supporter of the shortest-path concept must critizise Black's move as rather irrational: Black avoids loss of a minor piece and subsequently loses her Queen; similarly, a Queen-saving move is likely to delay the final mate longer than the move chosen in the actual game.

A similar conflict is shown by Roycroft's example (ibid. p. 102). The position is WKb8, WQe5, WRb5, BKg6, BQf4, and the plausible move for Black is Qf8+ avoiding the exchange of Queens and going for perpetual check, whereas the data base showed that the exchange of the Queens would provide a longer delay of the mate.

Black's behaviour in the cited examples is not an instance of weak play or an exception. On the contrary, it is the very practice in chess and can be observed frequently. The basic idea behind this kind of behaviour is: the player reliably <u>knows</u> that a certain move or outcome is bad for him or her and therefore avoids it without bothering how bad other alternatives may be or what the opponent may do next. In the example of Diagram 1, Black was quite certain to lose if she could not recapture the Rook, so she recaptured. The point is that such behaviour is quite rational and therefore optimal too. What the player optimizes, of course, is not the number of moves (longest delay or shortest achievement), but the cognitive effort of arriving at a reliable decision. In this sense she does use a shortest-path criterion as well, but instead of a shortest path to mate it is a shortest-path to the nearest achievable position of assured evaluation.

A further example from Harley (1970, p. 4) (see Diagram 2) may illustrate this principle of least effort:



White to move
DIAGRAM 2

In a real game many would be tempted to play 1. Ra8+ (or Rh8+), then exchange the parrying Knight and win the KRK-endgame. But the position also is mate in two by 1. Rhg7 Zugzwang. Of course, the longer and more straightforward solution is "easier", and it is also easy to see why: the longer solution is a straight variation without any branching (up to the subgoal KRK), whereas for the short solution no fewer than two King and eight Knight moves have to be examined - quite apart from the shorter solution being harder to find.

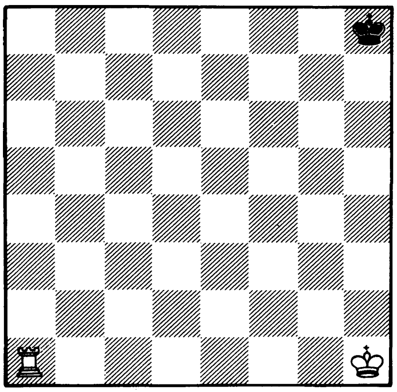
THE PRINCIPLE OF LEAST EFFORT

It must clearly be recognized that least-effort optimality depends on the problem solver's specific knowledge. The player's motto is "Reach (or avoid) the next possible subgoal" rather than "Reach (or avoid) mate". Different degrees of knowledge yield different subgoals and hence different search trees and choices. Therefore Roycroft's 1985 proposals are a step in the right direction. Still, his suggestions to determine general guidelines seems wrong to me. Instead, I feel that a flexible and enriched use of chess knowledge is to be preferred for the notion of optimality as may be indicated by the discussion below.

Let us take an example. Consider a KFKG endgame, F representing material stronger than or equal to G. Roycroft's proposal of preferring the stronger side to capture is convincing when direct mating (i.e., mating with neither F nor G reduced) is not possible or rather distant or difficult to foresee. But when the stronger side clearly has a choice between mating quickly and capturing, the latter looks unnatural. Refining the general rules in this way will lead to specifying exceptions and case distinctions. Thus, new rules may emerge such as "If F is a Queen or a Rook, then mating is to be preferred", "If Pawns are involved, promotion is to be preferred if ..." or "If feature X is present, then prefer outcome A, else B, except when ...".

If one admits this line of reasoning, it seems preferable first to apply specific endgame knowledge to the specific domain under investigation and only then, if desirable and possible, to try and extract general rules.

More generally speaking, rules need not be confined to captures and promotions only. Extending them is perhaps best illustrated by an example referring to data-base knowledge (see Diagram 3).



White to move
DIAGRAM 3

Capablanca (1921), having discussed this position, proposed 1. Ra7 leading to mate in 10 moves. Friedel (1985 b) classifies this move as an "error", since the data base showed 1. Rg1 to lead to mate in 9 moves only. No doubt this solution is shorter, but why should it be "better", and why should Capablanca's move be an "error", noting that Capablanca did not claim to give the shortest way to mate? I submit that there are excellent reasons for Capablanca's choice of 1. Ra7. In the diagrammed KRK position, there are at least two schemes of impelling the lone King (cf. Seidel, 1986); one of these is effected on the 8th rank (after 1. Ra7), the second on the h file (after 1. Rg1). Yet, the first scheme is simpler in structure and more familiar, whereas the second one requires special attention because of the danger of stalemate when the black King is in the corner. Therefore, on the principle of least effort (or cognitive economy), 1. Ra7 is preferable to 1. Rg1.

THE LEXICAL APPROACH

So far discussion of optimality has ignored the drawn case, where the shortest-path criterion completely fails: can any side have an interest in speeding up or delaying the final decision? It could be claimed that the weaker side has an interest to eliminate the danger of losing and therefore should speed up the decision, whereas the stronger side may have an interest in prolonging the agony, hoping that its opponent can be tempted to blunder into a suicidal move. But when the game-theoretical value of a position is drawn, no stronger or weaker side can be said to exist. In order to distinguish the interest of both sides, extraneous knowledge should be introduced flexibly. It seems to follow that, here again, Roycroft's proposed inflexible rules will be difficult to apply: there are, for instance, many cases where the draw can be reached without captures or other drastic events. I propose that the problem can be solved by exploiting knowledge of the specific endgame, for instance by introducing patterns of provable draws as possible subgoals.

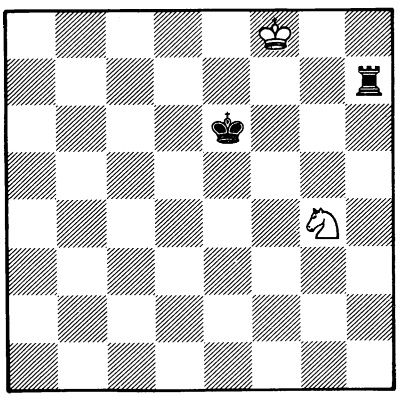
There is a further argument in favour of introducing more specific knowledge, especially into data-base construction. Data-base solutions of more complicated endgame positions have turned out to be very difficult to understand; for example, solutions of KBBKN positions have been qualified as "mysterious" even by distinguished experts (Friedel, 1985 a). One way out of this unsatisfactory quandary may be found by the following procedure. First, a set of plausible subgoals is specified, proven to be conditions for winning - proofs may be furnished by theory or data-base results. Next, the data base is constructed (or reconstructed) by working back from the subgoals rather than from a final mate or a reduced-material position. The logic design of this procedure is, of course, subtler than shortest-path construction and will depend on the characteristics of the subgoals, viz.

on whether they are merely sufficient or necessary, or both necessary and sufficient to reach the final result.

Extending the above, I should like to draw the conclusion that a lexical approach to the definition of optimal play should replace previous definitions of this notion. The lexical approach implies that the constructor of any endgame theory, whether or not derived from a data base, should first specify a list of subgoals he considers to be valid, hence reliable, on the domain in question. If desired, he may, additionally, define specific preferences, such as "direct mates in up to three moves shall be preferred to captures" or "transition to the specified endgame X shall be preferred to that to the specified endgame Y". Such a list of accepted doctrines and defined preferences may be quite short or, by contrast, as extensive as needed to embrace a theoretical grasp of the endgame in question; at any rate it should be designed for the specific domain.

It should be noted that by the lexical approach the shortest-path criterion is not ruled out per se; the approach merely determines the next subgoal and is in control of the selection of moves until such a subgoal is reached. It may well happen that the shortest-path principle even will bear the brunt of the decisional work; we simply remark that it should be subordinate to knowledge of subgoals.

Furthermore, relativation of the shortest-path rule, of course, does not exclude the case that a shorter solution is the better one (by being shorter). The following position (see Diagram 4), from a game of Steinitz' (Euwe, 1970), exhibits an instance of the optimality of the shortest path even under my reservation.



Black to move
DIAGRAM 4

Black continued 88. ... Rh4, forcing the Knight to move and catching it 8 moves later. It has recently been discovered (Hiltner, 1985) that 88. ... Rh3 would have won the Knight at once by Zugzwang. Since in all likelihood the 8-move sequence actually played (which implies several variations) was not available as previously formulated lexical knowledge and therefore would have required some analysis, the shortest-path solution by Rh3 is indeed the optimal one, both for economy of path and for cognitive economy.

A SHORTEST PATH MIGHT BE MYSTERIOUS

To anyone accepting my arguments, the claim of the shortest-path principle to provide a unique optimal solution must be rejected. This rejection is not, of course, valid when the principle is applied to the legitimate purpose of cataloguing shortest-path solutions. Still, there remains a severe objection against an exclusive shortest-path or "ulti-mate" orientation in endgame analyses as has, e.g., been advocated by Van Bergen (1985): the strict shortest-path criterion obscures the principles rather than revealing them. One of Van Bergen's examples may demonstrate this. The position is WKc7 WPc5 BKc4 and, according to the author, the ulti-mate oriented data base prescribes 1. Kd6. Since there is no interest at all to consider the length of the KQK ending following the Pawn's promotion, my view of chess theory and a practical player both will regard 1. c6 as the natural move, being the quickest way to the subgoal of queening. Offering the "optimal" move 1. Kd6 probably will have no other effect than increasing the "mysteriousness" of data-base results.

REFERENCES

Bergen, A. van (1985). An Ulti-Mate Look at the KPK Data Base. ICCA Journal, Vol. 8, No. 4, pp. 216-218.

Bramer, M.A. (1980). An Optimal Algorithm for King and Pawn against King Using Pattern Knowledge. Advances in Computer Chess 2 (ed. M.R.B. Clarke), pp. 82-96. Edinburgh University Press, Edinburgh.

Capablanca, J.R. (1921). Chess Fundamentals. G. Bell and Sons, Ltd., London.

Euwe, M. (1970). Feldherrnkunst im Schach. (Partie Nr. 21, pp. 62-68). Walter de Gruyter & Co., Berlin.

Friedel, F. (1985 a). Das perfekte Endspiel (2). Computer-Schach & Spiele, Vol. 2, No. 4, pp. 22-25.

Friedel, F. (1985 b). Fehler bei Capablanca. Computer-Schach & Spiele, Vol. 2, No. 5, pp. 29.

Harley, B. (1931). Mate in Two Moves. Reprinted by Dover Publications Inc. (1970), New York.

Hiltner, A. (1985). Aus Leserbriefen. Zwei Entdeckungen. Deutsche Schachzeitung, Vol. 134, No. 11, pp. 386-387.

Plesse, H.-J. (1985). Kombinationen aus aller Welt. Deutsche Schachzeitung, Vol. 134, No. 4, p. 132 (V).

Roycroft, A.J. (1985). Chess-Endgame Data-Base 'Oracles': Necessary and Desirable Features. ICCA Journal, Vol. 8, No. 2, pp. 100-104.

Seidel, R. (1986). Deriving Correct Pattern Descriptions and Rules for the KRK Endgame by Deductive Methods. Advances in Computer Chess 4 (ed. D. Beal), pp. 19-36. Pergamon Press, Oxford.

Photo by L. Lindner

REINER SEIDEL IN MILAN,
Optimality can be counted on the fingers of one hand ...