NOTES
CHESS ENDGAME NEWS
G.M'C. Haworth

Reading, UK

The recent focus by Newborn and Hyatt (2014) on the increasing ability of chess engines to play endgames is most welcome. They revisited a 16-position test set TS1 derived from Fine (1941) and demonstrated that the seminal chess engine CRAFTY (Hyatt, 2015) backed only by 5-man ‘EGT’ endgame-tables (Nalimov et al, 2000) could now handle it with ease. They therefore considered a second 16-position test set TS2 as a stiffer benchmark challenge for CRAFTY other chess engines.

The following engines are defined here:
- \( M_{l,i} \): filters moves in sub-\( k \)-man (‘s\( k \)m’) positions by minimaxing on Depth to Mate (‘DTM’),
- \( F \): filters moves by minimaxing on the DTF depths defined by FINALGEN (Romero, 2012),
- \( C_{l} \): filters by minimaxing on ‘DTC’ Depth to Conversion and using s6m DTC EGTs,
- \( Z_{l} \): filters by minimaxing on ‘DTZ’ Depth to Zeroing of the ply-count, using s7m DTZ\( _{s60} \) EGTs,\(^3\)
- \( E = M \cdot F \cdot C_{l} \cdot Z_{l} \): filters by deploying engines \( M_{l} \cdot F \cdot C_{l} \) and \( Z_{l} \) in turn,
- \( H \): Hyatt’s CRAFTY, unassisted by EGT support,
- \( HM_{l} \): the engine used by Newborn and Hyatt (2014),
- \( X \): the author’s FRITZ14 engine, analysing at 3mins/position, and
- \( EH \): notional engine, EGT-based but supported by CRAFTY as needed, whose performance is defined here.

FINALGEN (Romero, 2012) provides depth to mate and/or winning pawn-conversion for positions with at most one piece per side and sufficiently limited pawn mobility. Because it cannot contemplate, e.g., endgame KQKQP, DTF depths can be greater than DTM depths. FINALGEN builds its EGTs in single-thread mode and does not call on non-FINALGEN EGTs. Engine \( E \) considers the four depth metrics in ‘most distant first’ sequence, the most describable of the twenty-four sequences available! The use of the DTF/CZ metrics does not affect ‘moves to mate’ but can isolate a unique optimal move when DTM alone does not. Table 3 gives example positions and moves, also illustrating the sort of unnatural move-choices that (non-DTM) metric arithmetic can dictate.

\( EH \) and \( HM_{l} \) can choose different moves but where the position is beyond all available EGT-based machines, \( EH \) is effectively engine \( H \) and, here, plays the move attributed to CRAFTY by Newborn and Hyatt (2014). The objectives of deploying engine \( EH \) on test sets TS1 and TS2 were to:
- exercise FINALGEN and the Lomonosov 7m DTM EGTs (MVL, 2015) where possible,
- examine to what extent each ‘\( EH \) element’ contributed in finding a best line from the test positions,
- compare the move-choices and ‘moves to mate’ of \( HM_{l} \) and \( EH \),
- examine the uniqueness and optimality of the moves available,
- consider what the characteristics might be of good positions in a notional test set TS3.

Table 1 details the positions of test sets TS1 and TS2, and indicates the performance of engines \( HM_{l} \) and \( EH \) on them. Note that these are mainly wtm wins except for a wtm draw (TS2.07), a btm draw (TS2.02) and three btm wins for Black (TS1.13 and TS2.04/05). It should also be noted that TS1.15 = TS2.13.

Table 2’s row \( a \) indicates the initial number of men for each position: row \( b \) provides a DEEP FRITZ14 3-minute evaluation of the initial position. Row \( c \) gives the number of positions which are beyond the scope of engine \( E \), with row \( d \) giving the first position checked by \( EH \) and row \( e \) giving the number of men at that point. Rows \( f-i \) indicate the number of positions where, respectively, FINALGEN, 7-man (7m), 6m and 5m DTM EGTs are the first endgame tools used within \( EH \): this data is also illustrated graphically in more detail in Figure 1. Row \( k \) indicates the first position at which engines \( EH \) and \( HM_{l} \) differ, with row \( l \) indicating the nature of the ‘suboptimality’ from \( EH \)’s

---

1 The University of Reading, Berkshire, UK, RG6 6AH. email: guy.haworth@bnc.oxon.org.
2 Bourzutschky and Konoval computed all 6m DTC EGTs (Haworth, 2013) but these are not publicly available.
3 The use of De Man’s s7m DTZ\( _{s60} \) EGTs (CPW, 2013) is valid: the FIDE 50m draw-rule does not become relevant here.
4 Further, \( C_{l} \) and \( Z_{l} \) prefer/deter a change of force or pawn-push even if there is no EGT, q.v., Table 3, #09.
Table 1. Key data for the positions of test sets TS1 and TS2.

<table>
<thead>
<tr>
<th>TS#</th>
<th>Identity</th>
<th>#m-w-b</th>
<th>Material</th>
<th>FEN</th>
<th>Val. 'Best' move</th>
<th>dim</th>
<th>Mate, ply</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>Fine 25</td>
<td>5-3-2</td>
<td>KPKKPP</td>
<td>6K1/7p/5P1/8/8/7P/8 w - - 0 1</td>
<td>1.0 Kg5</td>
<td>35 35 35 0</td>
<td></td>
</tr>
<tr>
<td>1.02</td>
<td>Fine 26</td>
<td>5-3-2</td>
<td>KPKKPP</td>
<td>8/2K5p1/4P1K5/8/8/8 w - - 0 1</td>
<td>1.0 Kg5</td>
<td>33 31 33 -2</td>
<td></td>
</tr>
<tr>
<td>1.03</td>
<td>Fine 29</td>
<td>5-3-2</td>
<td>KPKKPP</td>
<td>4K/3P4/2P2P/2K3/8/8/8 w - - 0 1</td>
<td>1.0 Kg5</td>
<td>25 29 25 4</td>
<td></td>
</tr>
<tr>
<td>1.04</td>
<td>Fine 42</td>
<td>5-3-2</td>
<td>KPKKPP</td>
<td>8K/2/8K1/2P1/2P2P w - - 0 1</td>
<td>1.0 Kg4</td>
<td>33 47 31 16</td>
<td></td>
</tr>
<tr>
<td>1.05</td>
<td>Fine 51</td>
<td>6-3-3</td>
<td>KPKKPP</td>
<td>7K1/8P1P3/2P3P1/3K1P3/8/8 w - - 0 1</td>
<td>1.0 Kg5</td>
<td>37 23 37 -14</td>
<td></td>
</tr>
<tr>
<td>1.06</td>
<td>Fine 53</td>
<td>6-3-3</td>
<td>KPKKPP</td>
<td>8K1P2/2P3K3/2P2P w - - 0 1</td>
<td>1.0 Kg4</td>
<td>47 37 47 -1</td>
<td></td>
</tr>
<tr>
<td>1.07</td>
<td>Fine 57</td>
<td>7-4-3</td>
<td>KP3/K3PP</td>
<td>8/8P/p3K3P2K2P3/2P w - - 0 1</td>
<td>1.0 Kg5</td>
<td>43 41 47 -16</td>
<td></td>
</tr>
<tr>
<td>1.08</td>
<td>Fine 61</td>
<td>10-5-5</td>
<td>KP1/KP(4)P1</td>
<td>8pp/K5/P3P3/8/8K8/8 w - - 0 1</td>
<td>1.0 Kg6</td>
<td>39 33 6</td>
<td></td>
</tr>
<tr>
<td>1.09</td>
<td>Fine 66</td>
<td>8-4-4</td>
<td>KP(4)P(3)</td>
<td>8/1k3/p3/8/K4/7/P5/8P1/8 w - - 0 1</td>
<td>1.0 Kg6</td>
<td>41 40 1</td>
<td></td>
</tr>
<tr>
<td>1.10</td>
<td>Fine 67</td>
<td>8-4-4</td>
<td>KP(4)P(3)</td>
<td>8K1/5/k4/p1K1/2K1/P1/2P5/8 w - - 0 1</td>
<td>1.0 Kg4</td>
<td>53 50 0</td>
<td></td>
</tr>
<tr>
<td>1.11</td>
<td>Fine 70</td>
<td>9-5-4</td>
<td>KP(4)P(3)</td>
<td>8P3/7K1/2P1/2P1/P2P/2K8 w - - 0 1</td>
<td>1.0 Kg5</td>
<td>65 62 -5</td>
<td></td>
</tr>
<tr>
<td>1.12</td>
<td>Fine 76</td>
<td>11-6-5</td>
<td>KP(4)P(3)</td>
<td>8K6/8P1/3P4/2P1/P2P1/P2P1/K1P w - - 0 1</td>
<td>1.0 Kg4</td>
<td>67 53 14</td>
<td></td>
</tr>
<tr>
<td>1.13</td>
<td>Fine 80</td>
<td>14-7-7</td>
<td>KP1/K6(6)P6(p)</td>
<td>8/1K1P4/4K1P3/8/P1P/K1PK/PK w - - 0 1</td>
<td>0.1 Kg5</td>
<td>39 43 34</td>
<td></td>
</tr>
<tr>
<td>1.14</td>
<td>Fine 82</td>
<td>11-6-6</td>
<td>KP1/K5(5)P5</td>
<td>2K5/8P2K2/3P3/2P w - - 0 1</td>
<td>1.0 Kg7</td>
<td>57 37 20</td>
<td></td>
</tr>
<tr>
<td>1.15</td>
<td>Fine 90</td>
<td>14-7-7</td>
<td>KP1/K6(6)P6(p)</td>
<td>7K1/6P1/2P2P1/2P1/2P w - - 0 1</td>
<td>1.0 Kg4</td>
<td>57 45 12</td>
<td></td>
</tr>
<tr>
<td>1.16</td>
<td>Fine 100A</td>
<td>12-6-6</td>
<td>KP1/K5(5)P5</td>
<td>7K1/6P1/2P2P1/2P1/2P w - - 0 0 1</td>
<td>1.0 Kg2</td>
<td>47 35 12</td>
<td></td>
</tr>
<tr>
<td>2.01</td>
<td>CCE1 479</td>
<td>8-4-4</td>
<td>KP(3)P(3)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 1</td>
<td>1.0 Kg4</td>
<td>47 47 0</td>
<td></td>
</tr>
<tr>
<td>2.02</td>
<td>CCE1 491A</td>
<td>8-4-4</td>
<td>KP(3)P(3)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 1</td>
<td>1.0 Kg7</td>
<td>47 43 40</td>
<td></td>
</tr>
<tr>
<td>2.03</td>
<td>CCE1 530</td>
<td>8-4-4</td>
<td>KP(3)P(3)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 1</td>
<td>1.0 Kg7</td>
<td>83 73 10</td>
<td></td>
</tr>
<tr>
<td>2.04</td>
<td>CCE1 608</td>
<td>8-4-4</td>
<td>KP(3)P(3)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 1</td>
<td>0.1 Kg5</td>
<td>39 55 -16</td>
<td></td>
</tr>
<tr>
<td>2.05</td>
<td>CCE1 679</td>
<td>14-7-7</td>
<td>KP6(6)P6(p)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 0 1</td>
<td>0.1 Kg5</td>
<td>79 67 12</td>
<td></td>
</tr>
<tr>
<td>2.06</td>
<td>CCE1 680</td>
<td>16-8-8</td>
<td>KP7(K7)P7</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 1</td>
<td>1.0 Kg5</td>
<td>43 43 0</td>
<td></td>
</tr>
<tr>
<td>2.07</td>
<td>CCE1 765</td>
<td>11-6-5</td>
<td>KP1/K5(5)P5</td>
<td>8/1K6P/2P2/2P2/2P2/P1K w - - 0 1</td>
<td>1.0 Kg5</td>
<td>34 34 10</td>
<td></td>
</tr>
<tr>
<td>2.08</td>
<td>Lejene 2</td>
<td>7-4-3</td>
<td>KP3/K3PP</td>
<td>k7/P4/3P/4P/S1P2/2P2/2K w - - 0 0 1</td>
<td>1.0 Kg5</td>
<td>65 65 0</td>
<td></td>
</tr>
<tr>
<td>2.09</td>
<td>Lejene 4</td>
<td>8-4-4</td>
<td>KP3/K3PP</td>
<td>k7/P4/3P/4P/S1P2/2P2/2K w - - 0 0 1</td>
<td>1.0 Kg4</td>
<td>61 61 0</td>
<td></td>
</tr>
<tr>
<td>2.10</td>
<td>Lejene 5</td>
<td>7-4-3</td>
<td>KP3/K3PP</td>
<td>k7/P4/3P/4P/S1P2/2P2/2K w - - 0 0 1</td>
<td>1.0 Kg3</td>
<td>49 49 0</td>
<td></td>
</tr>
<tr>
<td>2.11</td>
<td>Lejene 6</td>
<td>7-4-3</td>
<td>KP3/K3PP</td>
<td>k7/P4/3P/4P/S1P2/2P2/2K w - - 0 0 1</td>
<td>1.0 Kg3</td>
<td>77 77 0</td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>Christmas tree</td>
<td>11-6-5</td>
<td>KP3/K3PP</td>
<td>c3/4K1/4P/4P1/2P1/2P1/2P1/2K w - - 0 0 1</td>
<td>1.0 Kg5</td>
<td>61 61 0</td>
<td></td>
</tr>
<tr>
<td>2.13</td>
<td>Pillsbury 1895</td>
<td>14-7-7</td>
<td>KP6(6)P6(p)</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 0 1</td>
<td>1.0 Kg5</td>
<td>42 42 0</td>
<td></td>
</tr>
<tr>
<td>2.14</td>
<td>Capablanca 1915</td>
<td>15-8-7</td>
<td>CP7(K7)P7</td>
<td>8/1K1P4/3K1/3P3K1/2K w - - 0 0 1</td>
<td>1.0 Kg4</td>
<td>55 49 6</td>
<td></td>
</tr>
<tr>
<td>2.15</td>
<td>Botvinnik 1944</td>
<td>12-6-6</td>
<td>KP5/K5(5)P5</td>
<td>8/1k3/K4/4P1/3P4K1/2P1/2P1/2K w - - 0 0 1</td>
<td>1.0 Kg5</td>
<td>65 65 -4</td>
<td></td>
</tr>
<tr>
<td>2.16</td>
<td>Botvinnik 1958</td>
<td>12-6-6</td>
<td>KP5(K5)P5</td>
<td>8/1K6PP/1P3/2P3/2K6/8 w - - 0 0 1</td>
<td>1.0 Kg4</td>
<td>69 69 2</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. The benchmarking of HMs lines by engine HE in the context of the move-profiles.
Some headlines from the results:

- The 4.5m and 2.6m tests TS.01-06 are solvable using accessible Nalimov s7m DTM EGTs,
- a further 4 tests, TS1.07 and TS2.08/10/11, are solvable using CVL (2015) 7m DTM EGTs,
- a further 9 tests, TS1.08-11 and TS2.02/03/07/09/12 are solvable if only FINALGEN is also used,
- the remaining 13 tests (TS1.12-16 and TS2.01/04/06/13-16) require the initial use of CRAFTY,
- across TS1/2, CRAFTY, FINALGEN, 7m, 6m and s6m DTM EGTs are EH’s lead evaluator as follows:
  - CRAFTY 78/181, FINALGEN 142/222, 7m DTM 82/97, 6m 46/112 and 5m 324/290 times, i.e.,
  - in % terms, CRAFTY 12/20, FINALGEN 21/25, 7m DTM 12/11, 6m 7/12 and 5m 48/32,
  - q.v., Figure 1 for a graphical representation of the breakdown per position,
- HM5 and EH played identically on 6 tests, TS1.06 and TS2.01/02/06/07/10.

**Figure 1. TS1/2: number of plies for which CRAFTY, FINALGEN or 7m/6m/s6m EGTs are EH’s lead evaluator.**

**Figure 2. TS1/2: distribution of the difference of CRAFTY’s and EH’s ‘#moves to mate’.**

Figure 2 shows the frequency of mate-length differences other than zero, of which there are thirteen. HM5 in fact only takes 21 more moves across the 32 positions of TS1/2 than EH: the four outliers are TS1.05/TS1.07/TS2.04 where HM5 concedes for the defense, and TS1.14 where HM5 takes 10 moves more to mate. From TS1.05, a total of 7 moves are conceded with moves 1b (4m), 3b (2m) and 4b (1m). From TS.07, a net total of 8 moves are conceded, see moves 1b/2w/2b/4b/6w/8w. From TS2.04, the three moves 10. Ke2 Ke4 11. Kd2 concede 4, 4 and then crucially 7 moves respectively in DTM terms. For TS1.14, White’s moves 12/13/14/16 concede 1/1/1/5 moves in DTM terms.

HM5 resolves all the TS1/2 tests whereas, for 13 positions, E is as useful as a Mark 1 Dalek in a stairwell and requires CRAFTY’s initial help. Although all 5m EGTs can be created in nine phases as ancillary, parallel threads of computation within CRAFTY’s hour, it would be interesting to see the performance of CRAFTY, completely unassisted by M5, both playing itself and with EH taking the other side.

---

5 The 9 phases are: 3-man (2 phases = 0/1 pawns), 4-man (3 phases = 0/1/2 pawns) and 5-man (4 phases = 0/1/2/3 pawns).

Phases 1-5 take ~30 seconds and phases 6-9 take ~30m each, times easily improved with more parallelism.
Considerations of the two key resources, space and time, suggest that the \textit{EH/HM} performance comparison is one of unlike ‘apples and pears’. In the latter days of man-machine contests, the use of 6-man EGTs was banned for these reasons. Engine \textit{E} inherits the unlimited space/time resources used to compute EGTs whereas \textit{HM} is using predetermined space/time and only one hour of real-time solving time. The use of ‘WDL’ win/draw/loss EGTS on ever-greater GBytes of low-latency SSD memory will facilitate access to EGTS deeper into the forward search. Figure 1 shows how far down the search-tree \textit{CRAF TY} would have to progress before invoking a ‘\textit{FINALGEN}’ to create some EGTS: 2 ply in TS2.01, 3 ply in TS1.12 but 44 ply in TS2.05/06.

There are at least two measures of the ‘difficulty’ of a position. One is the time taken to identify and adopt what is the best move but this will reduce as hardware evolves. A second, more hardware-independent measure, is the apparent relative merit of ‘the best move’ at various depths of search by one or more engines, though it is not yet clear how this can be turned into a single number even for one engine.

As a footnote, the criticality of the position and value \(\nu\) of ‘the move’ can be assessed by analyzing the TS1/2 positions with the other side to move (ostm). Verdicts after \(\bigcirc\) are not purely EGT-based but required some tree-search and evaluation:

\begin{itemize}
  \item \(\nu = 1\) point, win becomes a loss: TS2.01; \(\bigcirc\) TS1.12/15, TS2.06/13
  \item \(\nu = 1/2\) point, win becomes a draw: TS1.01/05-11, TS2.03/08-11; \(\bigcirc\) TS1.14, TS2.04-05/15-16
  \item \(\nu = 1/2\) point, draw becomes a loss: TS2.02
  \item \(\nu = 0\) point, result unchanged:
    \begin{itemize}
      \item - TS1.02-04, TS2.12. Note that TS1.03/04 and TS2.12 are type-BM zugs;\(^6\)
      \item - \(\bigcirc\) TS1.13/16, TS2.07/14
    \end{itemize}
\end{itemize}

Clearly, as chess engines search more deeply and therefore improve, the creation of challenging test sets becomes harder. Their purpose is primarily to test chess-engines’ search and evaluation rather than their use of pre- or even runtime-computed EGTS. Therefore, while the value of positions should be known, they should not be clearly decisive, have best opening moves which are quickly found or be in an EGT or in range of \textit{FINALGEN}. Only a few TS1/2 positions, including TS2.01/09/10/12, distinguish themselves in this regard today. The focus on pawns, especially those with restricted movement, and the initial exclusion of pieces is helpful to both \textit{FINALGEN} and chess-engine search, so the exclusive use of KP-endsgame is an onerous restriction but one which is fortunately unnecessary.

The Chess Study epitomizes the ‘hard to solve’ position and TS1/2 used 16 of these. Some other investigations of ‘anti-computer’, even ‘impossible to solve’, positions, have drawn entirely on the corpus of studies, currently represented without peer by van der Heijden’s \textit{HHoBV} (2010). Three notable articles are those by van der Heijden himself (2006, 2014) and Vlasák (2013). However, it should be said that many of their choices look more like ‘game’ than ‘endgame’ positions, one having as many as 22 men on the board. This suggests that there should be separate accolades for the most difficult \(m\)-man positions. Do those positions of most marginal advantage with the greatest metric-depth (Haworth, 2013a/b) provide the greatest challenge to chess-engines if they do not have access to the relevant endgame table?

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
# & TS move & FEN & E-DTM & E-DTF & E-DTC & E-DTZ & Notes \\
\hline
01 & 1.01 & 1w & 6k1/7p/5P1K1/8/8P/8 w - - 0 1 & Kg5 & = & Absolutely unique winning move \\
02 & 1.11 & 3w & 8k2/5p3/p3P/2P1P1/2K3/8K3 w - - 0 1 & Kd1 & = & Effectively unique: alternative Kb1 merely retracts move 2w \\
03 & 1.01 & 1b & 6k1/7p/5P1K1/8/8P/8 b - - 0 1 & Kf7 & = & The DTM EGT decides \\
04 & 1.01 & 10w & 5k5/2P5K1/8/8P/8 w - - 0 1 & h6 & h6 & = & The DTF EGT decides \\
05 & 1.01 & 2w & 2k7/5P2K1/8/8P/8 w - - 0 1 & Kf5 & Kf5 & = & The DTC EGT decides \\
06 & 2.08 & 7w & 7k5/5P3/2K3P2/5P8 w - - 0 1 & Ke5 & Ke5 & = & The DTZ EGT decides \\
07 & 1.12 & 14w & 8k4/3P1P1/8P3/8P8 w - - 0 1 & Kd3 & Kd3 & = & DTZ decides without EGT: immediate capture \(\Rightarrow: dtc = 1p\) \\
08 & 1.09 & 2w & 8k1/4P3K5P/7P5/8P1/8 w - - 0 1 & ? & Kd6 & Kd6 & = & DTZ decides without EGT: immediate P-push \(\Rightarrow: dtc = 1p\) \\
09 & 1.12 & 24w & 8k5/2/4Q4/4K/1P1/P7/8 w - - 0 1 & Kg7 & Kg7 & = & DTZ decidw without EGT: a4b4 and a4 are both rejected \\
10 & 1.14 & 15b & 7Q3/6P/6P1/8P1/8/6P w - - 0 1 & Ke5 & Ke5 & = & ... \(\text{dim} < \text{dtf}; \text{dim} = 8p \text{and} \text{dtf} = 16p\) \\
11 & 1.14 & 16w & 7Q3/6P/6P1/8P1/8/6P w - - 0 1 & bQb & bQb & = & ... and not CRAFTy 25. Kc6, reaching for the 5m EGTS \\
12 & 2.03 & 25w & 8k3/4P5K1/4P1/8P6/8 w - - 0 1 & cQc & cQc & = & DTZ-excluded. DTZ-a-optimal Kb6 requires Kc6(67)""" \\
13 & 1.12 & 20w & 6k1/4P3/4P1/4K1/1P1/P7/8 w - - 0 1 & f7 & f7 & = & ... and not CRAFTy. 20...Kc7 clearly prolongs the line \\
14 & 1.12 & 20b & 6k1/4P2/4P1/4K1/1P1/P7/8 w - - 0 1 & Kb8 & Kb8 & = & Unnatural. 20... Kc6 promotes a pawn quickly \\
15 & 1.09 & 3w & 8k5/2/4Q4/4K/1P1/P7/8 w - - 0 1 & Kb6 & Kb6 & = & Unnatural. 20...Kc7 clearly prolongs the line \\
16 & 1.15 & 17b & 8k5/3/4P5/2/2P/8P1/7P7K1/8 w - - 0 1 & Ke6 & Ke6 & = & Kc6 \(\Rightarrow: \text{dtf} = 24p \text{ though} \text{dim} = \sim 12p\). Kd6 is more natural \\
\hline
\end{tabular}
\caption{Table 3. Some illustrative positions and decisions taken by engine \textit{E}.}
\end{table}

\(^6\) A type-BM zug is one in which DTM is greater with the move than without it (Bleicher and Haworth, 2010)
Table 4. The studies cited by van der Heijden (2006, 2014) and Vlasák (2013).

Table 4 lists the 37 positions that van der Heijden and Vlasák chose. Authors are credited and serial numbers in the HHD8IV corpus are given. The status of the positions vis-à-vis sub-8-man EGTs and FINALGEN is indicated. Some of these positions may contribute in part to a future, hypothetical test set TS3.

The author lightly tested ‘DF14’ DEEP FRITZ 14, i.e., Horváth’s engine PANDIX (CPW, 2015) against the first, recommended move. The right-hand columns show its evaluation based on a 2-core, 3-minute run, the ‘A’ evaluation-difference to the next-best move, the move chosen and whether this agrees with the composition’s author or not. But DF14 is just one engine and it should be noted that different engines have different blind spots, can succeed or fail in finding ‘best moves’ and can certainly vary widely in their efficiency. The lesson is perhaps to stress-studies with a battery of significantly different engines. Vlasák used HARISS and HOU DINI, HARISS usually but not always being slower. Of his studies, the ones which defeated an engine or occupied it for more than one hour are V04-05, V12 and V16-18. Others, such as V03/07/08/10/14, are in the second, ‘useful engine-performance benchmark’ class while the rest are quickly solved by engines.

The author’s DF14 found the recommended first move in all but 8 of the 37 positions, but note that one of these, H17, is included only to emphasise that the study composer makes a falible appellant who can be tricked

7 HHD8IV indices for the 16-study subset of TS1/2 are: TS1.01-04 (47988, 7316, 1983, 18467), TS1.06-07 (1842, 20109), TS1.10-11 (3970, 4175), TS2.01 (15174), TS2.03 (15590), TS2.07-12 (51741, 66283, 18012, 31619, 66284, 7093).
whereas a computer engine presumes its opponent is no more fallible than itself. The full Saavedra study behind $H17$ also emphasises that the first move is not necessarily the crux of the study. Therefore, finding the first move is no guarantee that the engine will reproduce the intended and presumably correct solution, within a prescribed time or at all. CRAFTY did not reproduce the identical mainlines of all studies in TS1/2.

$H01$, now proved sound (Nunn, 2012) after years of debate, continues to defeat the best engines despite its short solution. $H02$ is only 7-man but also defeats DF14, $H18$ and $H19$ also provide a significant challenge.

Fortresses, perpetual check, zugzwangs (especially if engines’ ‘null move feature’ cannot be switched off) and the 50-move draw rule continue to be factors which pose difficulties for chess engines.

My thanks to the authors cited, and particularly to Harold van der Heijden and Emil Vlasák for their test sets of compositions, and to future readers who contribute suggestions for test set TS3.

References


---

8 This is a reflection on the concept of ‘correctness’. For $H17$, engines choose $5. \cdots \text{Rf3}$ postponing mate, but the swindle $5. \cdots \text{Rd4}$ hopes for $6. \text{c8Q??}$, a mere draw. Black’s most powerful piece is the white queen by the side of the board.

9 Perhaps more than 30% of studies have a flaw in their solution. Vlasák, e.g., notes that V12 is cooked by $1. \text{Rf7}$. HHDBIV notes these flaws where known and includes repaired versions of studies where available.