

The data-mining of studies database 'HHdbIV'

Article

Published Version

Haworth, G. ORCID: <https://orcid.org/0000-0001-9896-1448>,
Van der Heijden, H. and Bleicher, E. (2013) The data-mining of
studies database 'HHdbIV'. EG, 19 (191). pp. 27-30. ISSN
0012-7671 Available at <https://centaur.reading.ac.uk/30291/>

It is advisable to refer to the publisher's version if you intend to cite from the
work. See [Guidance on citing](#).

Publisher: ARVES

All outputs in CentAUR are protected by Intellectual Property Rights law,
including copyright law. Copyright and IPR is retained by the creators or other
copyright holders. Terms and conditions for use of this material are defined in
the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The Data-mining of Studies Database HHdbIV

BY GUY HAWORTH, HAROLD VAN DER HEIJDEN
& EIKO BLEICHER

In his recent Spotlight columns, Jarl Ulrichsen (2012a/b) made and later revisited a note that, to his surprise, it was still possible to find sub-7-man (s7m) positions in the studies of HHdbIV which signal cooks and/or duals. Particularly for those readers of EG who are similarly and in fact unnecessarily surprised, we explain here what exactly was done with the s7m-positions in creating HHdbIV.

The first author decided to examine the s7m positions in the mainlines of the studies in HHdbIII. The two goals, in priority order, were to examine the correctness and uniqueness of White's move, that is, to identify:

- those mainline positions with values incompatible with the stipulation of the study, and
- those mainline positions where there were alternative moves which pre-served value.

The set of win studies with s7m positions in the mainline was identified using CQL (Costeff, 2003) and the lines were converted into lists of positions using the utility PGN2FEN (Foden, 2010). These positions were re-grouped in line with the studies they came from and the now-redundant positions with at least 7 men were discarded. The same process also created a set of mainline positions from draw studies.

Both sets of positions were evaluated using Nalimov's DTM Endgame tables by Eiko Bleicher (2012), both with the side-to-move and, because the frequency of zugzwangs was also of interest (Haworth et al, 2011b), with the opposite side-to-move. The information was sent to Harold van der Heijden who, it transpired, was in the final stages of publishing HHdbIV. He asked for the 'HHdbIV studies' that were not in HHdbIII to be evaluated

as well. It was then clear that 3,068 studies contained between them some 18,741 s7m mainline positions whose value was incompatible with the stipulation of the study. In some cases, it was the stipulation that was wrong (i.e. data entry error).

Details of the data-mining exercises have been published (Bleicher et al, 2010; Haworth et al, 2011a). In a few months, Harold transformed almost all the 'wrong position value' information into '@-indicators' in HHdbIV with comments on a selection of moves as to what would have been the right move for White. The first author would therefore like to suggest that the attributions to cook-authors 'GH/EB', over 5,100 of them, should be changed in HHdbV to 'GH/EB/HH'. Harold also led in the publication of two articles in EG (Van der Heijden et al, 2010a/b) which focused on the chess aspects of some chosen cooked studies.

To emphasise the status of the work prior to the publication of HHdbIV, the first author did not examine the values of s7m positions in studies' sidelines. This was because no automatic method was known for determining whether a position was meant to be won or drawn. Further, there was no time before the publication of HHdbIV to process the data about alternative value-preserving moves as this requires a mix of difficult, automatable but not yet automated, technical assessment (Haworth and Rusz, 2012) as well as artistic, chess judgement.

It is worth putting the incidence of equi-optimal and suboptimal moves into context by distinguishing four types of mainline position. In 150,649 of 234,634 s7m positions, the value-preserving move is absolutely unique. In

59,409 positions, the DTM-optimal move is unique but there are DTM-suboptimal moves, and one of the latter was played in 8,167 positions. In 13,186 positions, there are only alternative equi-optimal moves. In 11,390 positions there are both equi-optimal moves and DTM-suboptimal moves, and one of the latter was played in 1,665 cases.

There were 320,579 DTM-suboptimal moves available, and 9,832 cases of one being chosen, indicating a dual of some sort if not a chessic or data entry error. These include missed mates in 1 (25), in 2 (67), in 3 (90), in 4 (129) and in 5 (172), many not remarked on to date. One might conjecture that the shallower the DTM-depth and the greater the DTM-concession, the greater is the error likely to be. Many DTM-suboptimal moves will in future be proved to be merely time-wasting

moves, allowing repetition of position or no better progress to the next mainline position.

There are 24,576 positions involving 44,227 equi-optimal moves where the downstream-convergence (or lack of it) of those equi-optimal moves should be examined.

When the technical assessment aspect has been automated in a relatively small production process (Haworth and Rusz, 2012), there will be an increased opportunity for the endgame community to make both technical comments and artistic judgements, particularly about the seriousness of duals.

The endgame scenario is likely to take one step forward soon as a set of, as yet unverified, DTM EGTs for 7-man chess has been created (Haworth, 2012; MVL, 2012). Lists of 7-man positions are to hand.

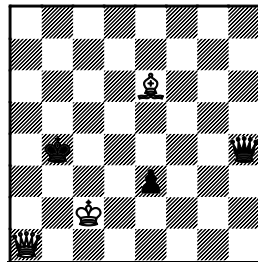
HHdbIV				DTM- conc./			DTM-opt.		
#	#	Author	GBR	Year	FEN	DTM	conc.	DTM	move
01	21861	Khatchaturov	+0041.00c6c8	1945	3k4/8/2KN4/b3B3/8/8/w--64	1	22	22.00	4. Bf6#
02	23282	Bron	+0052.12d3c1	1948	8/8/8/8/b3B3/2KN4/8/3c4w--410	1	22	22.00	10. Bf3#
03	70034	Borisenko	+4010.01c2b4	2003	4B3/2Q5/8/8/1k5q/4p3/2K5/8/w--2614	1	15	15.00	14. Qc3#
04	05118	Hatheway	+0041.00f6e8	1908	5k2/5B2/b1N2K2/8/8/8/8/w--64	2	24	12.00	4. Ne5
05	05414	Amehung	+0133.00b3a1	1909	8/8/3n4/8/2K5/8/8/7R/k7w--34	2	20	10.00	4. Kb3
06	05701	Troitzky	+0002.02d4f3	1910	8/8/7p/6pN7/k6N1/k6K1/8/w--169	3	20	6.67	9. Nf6#
07	00870	Cozio	+1300.01a2b8	1766	k7/p2Q4/1r6/K7/8/8/8/w--64	4	18	4.50	4. Qc8+
08	17800	Dedrie	+1300.01h3h1	1937	8/8/8/8/8/5Q1K/4r2p/6k1w--22	5	32	6.40	2. Qze2
09	69962	Borisenko	+0014.00b6b8	2003	1k6/8/8/1NK2B2/8/8/n7/8/w--3016	6	31	5.17	16. Kb6
10	03841	Meyer	+0002.02f6h6	1897	8/7k/5K2/4p3/4p2N/4B3/8/8/w--127	7	28	4.00	7. Ne5
11	19331	Dedrie	+1060.00f5c2	1939	8/b7/8/8/1Q6/1b6/2k1K3/8/w--64	8	43	5.38	4. Qd2
12	21450	unknown	+0116.00g3c5	1944	8/4n3/4R1n1/2k5/8/8/B4K2/8/w--43	30	108	3.60	3. Rf1*
13	20833	Szule & Kopac	+0134.00d4e6	1942	8/8/4k3/1N6/n2K4/R7/2b5/8/w--64	50	87	1.74	4. Ra1*

Most of the duals spotted (see Table) are boring from an artistic point of view, but there are some exceptions:

(D.1) Intended: 1.Qc3+ Kb5 2.Bd7+ Kb6 3.Qc6+ Ka7 4.Qc5+ Kb7 5.Bc6+ Kc7 6.Bb5+ Kb8 7.Qd6+ Kb7 8.Bc6+ Kb6 9.Bd5+ Kb5 10.Qc6+ Ka5 11.Qc7+ Kb5 12.Bc6+ Kc5 13.Be8+ Kb4 14.Qd6+ Kc4 15.Qc6+ Kd4 16.Qa4+, overlooking 14.Qc3 mate!

Not all sub- (and equi-) optimal moves spoil an endgame study. The following illustrative example **(D.2)** has been used before, but it does no harm to examine it again:

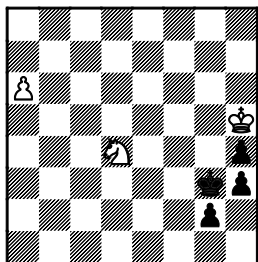
D.1. I. Borisenko
3rd prize *Narodnaya Tribuna* 2003



c2b4 4010.01 3/3 Win

D.2. H. Aloni

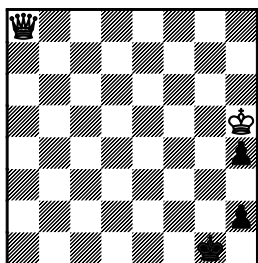
2nd commendation *Szachy* 1960



h5g3 0001.13 3/4 Win

(D.2) We concentrate on the main line: **1.a7 g1Q 2.Se2+ Kf2 3.Sxg1 h2 4.Sh3+ Kg3 5.Sf2 Kxf2 6.a8Q Kg1 7.Qa1+ Kg2 8.Qb2+ Kg1 9.Kxh4 h1Q+ 10.Kg3** wins.

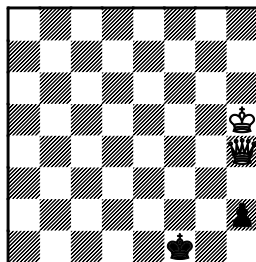
This is the position after move 6:



The EGTB indicates three winning moves: **7.Qa1+** (solution) with a DTM of 14 moves, **7.Qa7+** DTM: 14 and **7.Qg8+** DTM: 13 moves. Many people now jump to conclusions: **7.Qg8+** must be a cook as the EGTB indicates a shorter DTM.

Let's examine the optimal moves: **7.Qg8+ Kf2 8.Qd5 Kg1 9.Qg5+ Kf1 10.Qxh4**, and now we are in a very interesting critical position: (see diagram next column)

Now the EGTB indicates that **10...Ke2!**? is the optimal move for Black with a DTM of 9 moves. Of course, an o.t.b. player would never consider this move since **11.Qxh2** is an immediate win. The more natural move is **10...Kg2** (DTM 8) **11.Qe4+ Kg1 12.Qe1+ Kg2 13.Kg4 h1Q 14.Qe2+ Kg1 15.Kg3** This final position



(at move 15) is almost identical to the main line (at move 10!) and has the same idea. So **8.Qg8+** is merely a time wasting dual. That seems to be strange, because Black did not play DTM-optimally in this line. The explanation is that, in the solution, Black plays the natural **8...Kg1** (DTM 8), while **8...Kg3** (DTM 12) is optimal. But this also is a weak move to an o.t.b. player who would quickly find **9.Qb7** followed by **10.Qh1** leaving Black without any chance.

In conclusion, one should keep in mind that a DTM-optimal defence may be a very stupid move for a player or endgame study composer. Every claim based on an EGTB needs to be thoroughly examined, even if the EGTB indicates that the alternative has a shorter DTM.

References

Bleicher, E. (2012). <http://www.k4it.de/index.php?topic=egtb&lang=en>. Nalimov-EGT query service.

Bleicher, E., Haworth, G.M[°]C. and Van der Heijden, H.M.J.F. (2010). Data-mining Chess Databases. *IC-GA Journal*, Vol. 33, No. 4, pp. 212-214. <http://centaur.reading.ac.uk/17497>.

Costeff, G. and Stiller, L. (2003). <http://www.rbnn.com/cql/> CQL: Chess Query Language.

Foden, T. (2010). <http://www.pgn2fen.com-about.com/> The PGN2FEN v1.0.4 format-conversion utility.

Haworth, G.M[°]C. (2012). Chess Endgame News. *IC-GA Journal*, Vol. 35, No. 1, pp. 90-3. <http://centaur.reading.ac.uk/29422>.

Haworth, G.M[°]C., Bleicher, E. and Van der Heijden, H.M.J.F. (2011a). Uniqueness in Chess Studies. *IC-GA Journal*, Vol. 34, No. 1, pp. 22-24. <http://centaur.reading.ac.uk/19484>.

- Haworth, G.M^cC. and Rusz, Árpád (2012). Position Criticality in Chess Endgames. ACG13: Advances in Computer Games. Tilburg, the Netherlands. LNCS 7168 pp. 244-257. <http://centaur.reading.ac.uk/23799>.
- Haworth, G.M^cC., Van der Heijden, H.M.J.F. and Bleicher, E. (2011b). Zugzwangs in Chess Studies. *ICGA Journal*, Vol. 34, No. 2, pp. 82-88. <http://centaur.reading.ac.uk/23047>.
- MVL team (2012). <https://plus.google.com/100454521496393505718/posts>.
- Van der Heijden, H., Bleicher, E., Haworth, G.M^cC. (2010a). Endgame table testing of studies, I. **EG** 16.180 pp114-119. <http://centaur.reading.-ac.uk/4628>.
- Van der Heijden, H., Bleicher, E., Haworth, G.M^cC. (2010b). Endgame table testing of studies, II. **EG** 16.181 pp. 163-169. <http://centaur.reading.-ac.uk/5908>.
- Ulrichsen, J. (2012a/b). Spotlight. **EG** 189, pp. 203-4 and **EG** 190, p305.