Chess Endgame News

Article

Published Version

NOTES

CHESS ENDGAME NEWS

G.M^c C. Haworth[1]

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_{k-1}: 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_5: filters by minimaxing on 'DTC' Depth to Conversion and using s6m DTC EGTs,[2]
- Z_6: filters by minimaxing on 'DTZ' Depth to Zeroing of the ply-count, using s7m DTZ_{50}' EGTs,[3,4]
- $E \equiv M_7FC_5Z_6$: filters by deploying engines M_7, *F*, C_5 and Z_6 in turn,
- *H*: Hyatt's CRAFTY, unassisted by EGT support,
- HM_5: 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 KQQKP, 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/C/Z 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_5 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_5 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_5 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 s6m 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_5 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_{50}' EGTs (CPW, 2013) is valid: the FIDE 50m draw-rule does not become relevant here.

[4] Further, C_n and Z_n prefer/defer a change of force or pawn-push even if there is no EGT, q.v., Table 3, #09.

point of view. A further 3-minute evaluation by FRITZ10 at that point as given in row *m* makes it clear that, apart from position TS2.11, any engine would likely secure a result from those positions. Data beyond row *m* indicates the count of men for each subsequent position in the HM_5 line provided. Haworth (2015) provides supporting data, including extensions of the tables, pgn files and annotated lines of play.

TS#	Identity	#m	w-b	Material	FEN	Val.	'Best' move	*dtm*	Mate, ply HE	EH	Δ
1.01	Fine 25	5	3-2	KPPKP	6k1/7p/5P1K/8/8/8/7P/8 w - - 0 1	1-0	Kg5	35	35	35	0
1.02	Fine 26	5	3-2	KPPKP	8/2k5/p1P5/P1K5/8/8/8/8 w - - 0 1	1-0	Kd5	33	31	33	-2
1.03	Fine 29	5	3-2	KPPKP	4k3/4Pp2/5P2/4K3/8/8/8/8 w - - 0 1	1-0	Kf5	25	29	25	4
1.04	Fine 42	5	3-2	KPPKP	8/5p2/8/4K1P1/5Pk1/8/8/8 w - - 0 1	1-0	Ke4	33	47	31	16
1.05	Fine 51	6	3-3	KPPKPP	8/8/2pp3k/8/1P1P3K/8/8/8 w - - 0 1	1-0	d5	37	23	37	-14
1.06	Fine 53	6	3-3	KPPKPP	8/8/3pkp2/8/8/3PK3/5P2/8 w - - 0 1	1-0	Ke4	47	47	47	0
1.07	Fine 58	7	4-3	KP(3)KPP	8/8/2ppk3/8/2PPK3/2P5/8/8 w - - 0 1	1-0	d5+	43	31	47	-16
1.08	Fine 61	10	5-5	KP(4)KP(4)	8/ppp5/8/PPP3kp/8/6KP/8/8 w - - 0 1	1-0	b6	?	39	33	6
1.09	Fine 66	8	4-4	KP(3)KP(3)	8/1k3ppp/8/3K4/7P/5PP1/8/8 w - - 0 1	1-0	Kd6	?	41	41	0
1.10	Fine 67	8	4-4	KP(3)KP(3)	8/2p5/3k4/1p1p1K2/8/1P1P4/2P5/8 w - - 0 1	1-0	b4	?	53	53	0
1.11	Fine 70	9	5-4	KP(4)KP(3)	8/k7/3p4/p2P1p2/P2P1P2/8/8/K7 w - - 0 1	1-0	Kb1	?	63	65	-2
1.12	Fine 76	11	6-5	KP(5)KP(4)	8/8/p6p/1p3kp1/1P6/P4PKP/5P2/8 w - - 0 1	1-0	f4	?	67	53	14
1.13	Fine 80	14	7-7	KP(6)KP(6)	8/8/1ppk4/p4pp1/P1PP2p1/2P1K1P1/7P/8 b - - 0 1	0-1	b5	?	43	39	4
1.14	Fine 82	12	6-6	KP(5)KP(5)	8/pp5p/8/PP2k3/2P2pp1/3K4/6PP/8 w - - 0 1	1-0	c5	?	57	37	20
1.15	Fine 90	14	7-7	KP(6)KP(6)	8/7p/2k1Pp2/pp1p2p1/3P2P1/4P3/P3K2P/8 w - - 0 1	1-0	e4	?	57	45	12
1.16	Fine 100A	12	6-6	KP(5)KP(5)	8/6p1/3k1p2/2p2Pp1/2P1p1P1/1P4P1/4K3/8 w - - 0 1	1-0	Kf2	?	47	35	12
2.01	CCE4 479	8	4-4	KP(3)KP(3)	8/1p4kP/5pP1/3p4/8/4P3/7K/8 w - - 0 1	1-0	e4	?	47	47	0
2.02	CCE4 491a	8	4-4	KP(3)KP(3)	8/1pp5/3k3p/PP6/2P2K2/8/8/8 b - - 0 1	=	Kd7	=	=	=	=
2.03	CCE4 530	8	4-4	KP(3)KP(3)	2k2K2/8/pp6/2p5/2P5/PP6/8/8 w - - 0 1	1-0	a4	?	83	73	10
2.04	CCE4 608	10	5-5	KP(4)KP(4)	8/pp3p2/8/6kp/8/3K1PP1/PP6/8 b - - 0 1	0-1	f5	?	39	55	-16
2.05	CCE4 679	14	7-7	KP(6)KP(6)	8/pp2k1pp/2p5/2P1p3/2P1P2P/6P1/P7/2K5 b - - 0 1	0-1	g5	?	79	67	12
2.06	CCE4 680	16	8-8	KP(7)KP(7)	8/1p6/p1p5/P1Pp2pp/1P1P1p1k/5P1P/6PK/8 w - - 0 1	1-0	g3/g4	?	43	43	0
2.07	CCE4 765	11	6-5	KP(5)KP(4)	8/1k6/p4p2/2p2P2/p1P2P2/2P5/P1K5/8 w - - 0 1	=	Kc1	=	=	=	=
2.08	Lejeune 2	7	4-3	KP(3)KPP	k7/4p3/4p3/8/8/3P1P2/5P2/K7 w - - 0 1	1-0	Kb2	?	65	65	0
2.09	Lejeune 4	8	4-4	KP(3)KP(3)	k7/8/1p6/p1p5/2P4K/8/PP6/8 w - - 0 1	1-0	a4	?	61	61	0
2.10	Lejeune 5	7	4-3	KP(3)KPP	8/8/p7/8/1P6/7p/P4k1P/3K4 w - - 0 1	1-0	a3	?	49	49	0
2.11	Lejeune 6	7	4-3	KP(3)KPP	8/5p1p/8/6k1/8/6P1/5PP1/7K w - - 0 1	1-0	Kh2	?	77	77	0
2.12	Christmas Cracker	11	6-5	KP(5)KP(4)	3k4/3p4/3p4/p2P2p1/P2P2P1/3P4/3K4/8 w - - 0 1	1-0	Kd1	?	55	55	0
2.13	Pillsbury 1895	14	7-7	KP(6)KP(6)	8/7p/2k1Pp2/pp1p2p1/3P2P1/4P3/P3K2P/8 w - - 0 1	1-0	e4	?	57	45	12
2.14	Capablanca 1919	15	8-7	KP(7)KP(6)	8/5p2/2kp1p1p/p1p2P2/2P5/7P/PP3PP1/6K1 w - - 0 1	1-0	a4	?	55	49	6
2.15	Botvinnik 1944	12	6-6	KP(5)KP(5)	8/1p3k2/p4ppp/3P4/1P6/4K2P/1P4P1/8 w - - 0 1	1-0	g3/g4	?	61	65	-4
2.16	Botvinnik 1958	12	6-6	KP(5)KP(5)	8/4pk2/1p4p1/1P2p3/3pP2P/3K2P1/4P3/8 w - - 0 1	1-0	Kc4	?	71	69	2

Table 1. Key data for the positions of test sets TS1 and TS2.

		Test set TS1																Test set TS2															
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**#**	**#**	**#**	**#**	**#**	**#**	**#**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**#**	**#**	**#**	**#**	**#**	**#**	**#**
a	**# men initially**	5	5	5	5	6	6	7	10	8	8	9	11	14	12	14	12	8	8	8	10	14	16	11	7	8	7	7	11	14	15	12	12
b	**Initial evaluation**	10	17	18	19	7	10	9	8	8	6.0	13	6	##	7	1.3	3	7.1	3.4	4.4	##	##	10	1.6	2.7	7.5	6.2	1.2	13	4.0	4.9	2.6	2.3
c	**# pos. unchecked**	0	0	0	0	0	0	0	0	0	0	0	2	23	10	28	15	2	0	0	7	43	44	0	0	0	0	0	0	28	34	13	10
d	**First '*EH*- check'**	1w	1w	1w	1w	1w	1w	1w	1w	1w	1w	1w	2w	13w	6w	15w	8b	2w	1b	1w	5w	23w	—	1w	1w	1w	1w	1w	1w	15w	18w	7b	6w
e	**... # men**	5	5	5	5	6	6	7	10	8	8	9	10	8	9	9	9	8	8	8	8	9	—	11	7	8	7	7	11	9	9	8	10
	***EH* lines:**																																
f	**# 'FinalGen' pos.**	0	0	0	0	0	0	0	15	4	5	39	42	2	7	7	21	0	11	18	8	25	0	36	0	29	0	0	31	7	16	4	37
g	**# 7-man pos.**	0	0	0	0	0	0	2	4	1	10	11	10	15	21	8	0	5	1	11	1	0	0	0	7	2	14	32	10	8	0	5	1
h	**# 6-man pos.**	0	0	0	0	2	11	1	2	10	1	16	0	0	0	3	0	8	15	18	7	0	0	0	7	3	5	1	15	3	0	8	22
i	**# sub-6-man pos.**	36	34	26	32	36	37	45	13	27	38	0	0	0	0	0	0	33	5	27	33	0	0	0	52	28	31	45	0	0	0	36	0
j	**Mated by *EH***	18b	17b	13b	16b	19b	24b	24b	17b	21b	27b	33b	27b	21w	19b	23b	18b	24b	=	37b	29w	35w	22b	=	33b	31b	25b	39b	28b	23b	25b	33b	35b
	HM_5 lines:																																
k	**1st 'non-E' move**	1b	3b	3w	2w	1b	—	1b	6b	2w	1b	15w	2b	14b	7b	15b	9b	—	—	22b	5b	23w	—	—	1b	11b	—	8b	16b	15b	21w	10w	6b
l	**D(*EH*, HM_5)**	-2m	-1m	+2m	+3m	-4m	—	-4m	-2f	+1f	-1f	+2f	-1f	+2m	-2f	-1f	-4f	—	—	-3m	+1f	-2f	—	—	-1z	-1f	—	-1m	-1m	-1f	-3f	+1f	-1f
m	**Eval. at that pt.**	11	17	31	20	6.4	—	8.1	8.9	9.3	6.0	23	6.9	-¥	28	9.3	7.6	—	—	9.6	##	-14	—	—	2.7	9.9	—	5.1	21	9.3	+¥	6.2	6.5
# men on board in HM_5 lines	**16**	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1w	—	—	—	—	—	—	—	—	—	—
	15	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	2w	—	—	—	—	—	—	—	1w	—	—
	14	—	—	—	—	—	—	—	—	—	—	—	—	1b	—	1w	—	—	—	—	—	1b	3b	—	—	—	—	—	—	1w	11w	—	—
	13	—	—	—	—	—	—	—	—	—	—	—	—	3w	—	2w	—	—	—	—	—	5b	4b	—	—	—	—	—	—	2w	13b	—	—
	12	—	—	—	—	—	—	—	—	—	—	—	—	3b	1w	4b	1w	—	—	—	—	6w	5w	—	—	—	—	—	—	4b	15b	1w	1w
	11	—	—	—	—	—	—	—	—	—	—	—	1w	4b	3b	7b	3b	—	—	—	—	17b	5b	1w	—	—	—	—	1w	7b	17w	5w	4b
	10	—	—	—	—	—	—	—	1w	—	—	—	2w	9w	5b	9b	7w	—	—	—	1b	21w	19b	—	—	—	—	—	7b	9b	17b	5b	6w
	9	—	—	—	—	—	—	—	3w	—	—	1w	6b	12b	6w	15w	8b	—	—	—	4w	23w	—	—	—	—	—	—	15w	15w	18w	7w	21b
	8	—	—	—	—	—	—	—	4w	1w	1w	14b	16b	13w	7b	18w	13w	1w	1b	1w	4b	25w	—	—	—	1w	—	—	15b	18w	21w	7b	22w
	7	—	—	—	—	—	—	1w	8b	4b	13b	23b	21b	14w	11b	21b	14w	2w	5b	10w	9w	—	—	—	1w	15b	1w	1w	16b	21b	—	9b	22b
	6	—	—	—	—	1w	1w	2b	—	6b	14b	—	22w	16w	17w	22b	16b	4b	7w	15b	13w	—	—	—	4b	16b	8w	17w	21b	22b	—	17w	27w
	5	1w	1w	1w	1w	2b	6b	5b	—	9w	19b	—	23b	17w	—	—	17b	8b	7b	24b	16b	—	—	—	8w	18w	10b	17b	—	—	—	19b	—
	Mated by HM_5	18b	16b	15b	24b	12b	24b	16b	20b	21b	27b	32b	34b	23w	29b	29b	24b	24b	=	42b	21w	41w	22b	=	33b	31b	25b	39b	28b	29b	28b	31b	36b

Table 2. The benchmarking of HM_5 lines by engine *EH* in the context of the force-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 MVL (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,
- HM_5 and *EH* played identically on 6 tests, TS1.06 and TS2.01/02/06/07/10.

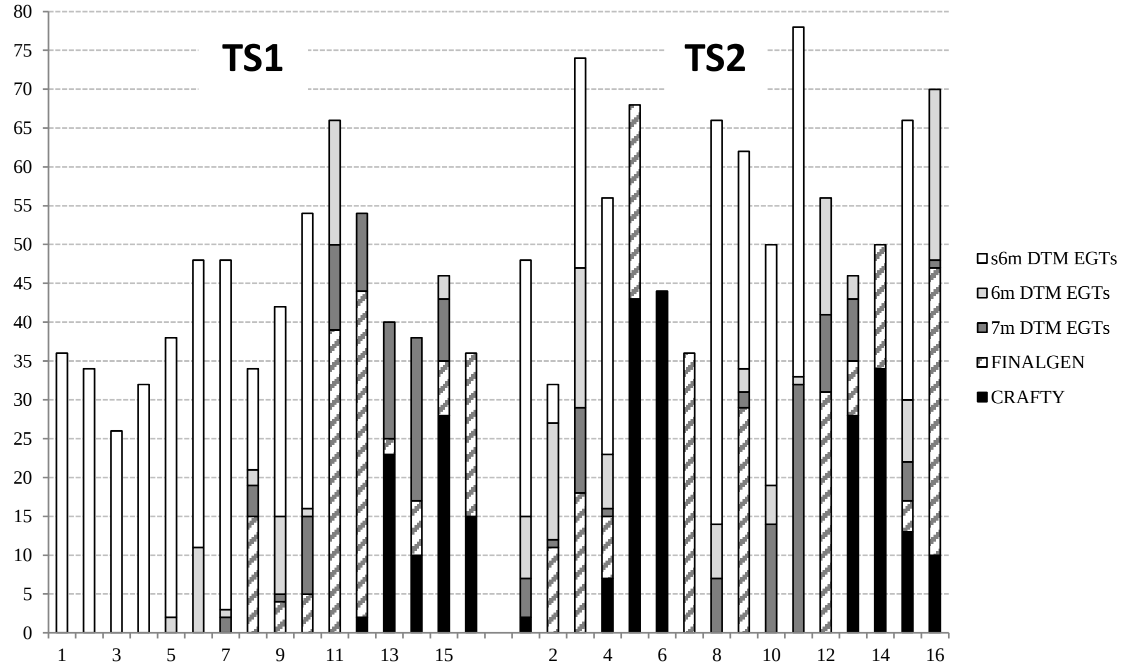

Figure 1. TS1/2: number of plies for which CRAFTY, FINALGEN or 7m/6m/s6m EGTs are *EH*'s lead evaluator.

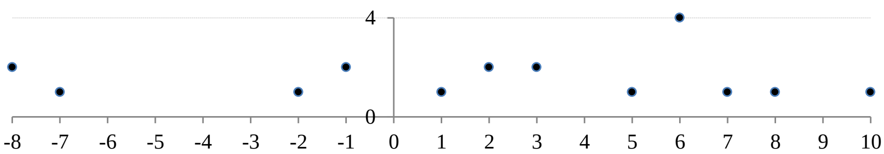

Figure 2. TS1/2: distribution of the difference of CRAFTY'S and *EH*'s '#moves to mate'.

Figure 2 shows the frequency of mate-length differences other than zero, of which there are thirteen. HM_5 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 HM_5 concedes for the defense, and TS1.14 where HM_5 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.

HM_5 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[5] as ancillary, parallel threads of computation within CRAFTY's hour, it would be interesting to see the performance of CRAFTY, completely unassisted by M_5, 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.

#	TS	move	FEN	E-DTM	E-DTF	E-DTC	E-DTZ	Notes
01	1.01	1w	6k1/7p/5P1K/8/8/8/7P/8 w - - 0 1	**Kg5′′′′**	—	—	—	Absolutely unique winning move
02	1.11	3w	8/2k5/3p4/p2P1p2/P2P1P2/8/8/2K5 w - - 0 1	?	Kd1′′′	—	—	Effectively unique: alternative Kb1 merely retracts move 2w
03	1.01	1b	6k1/7p/5P2/6K1/8/8/7P/8 b - - 0 1	**Kf7′′**	—	—	—	The DTM EGT decides
04	1.01	10w	8/5k2/5P2/6KP/8/8/8/8 w - - 0 1	h6′	**h6′′**	—	—	The DTF EGT decides
05	1.01	2w	5k2/7p/5P2/6K1/8/8/7P/8 w - - 0 1	Kf5′	Kf5′	**Kf5′′**	—	The DTC EGT decides
06	2.08	7w	8/4p3/8/4Pk2/2K2P2/8/5P2/8 w - - 0 1	Kc5′	Kc5′	?	**Kc5′′**	The DTZ EGT decides
07	1.12	14w	8/6k1/p7/1p3PKp/1P5P/P7/8/8 w - - 0 1	?	Kxh5′	**Kxh5′′**	—	DTC decides without EGT: immediate capture ⇒ *dtc* = 1p
08	1.09	2w	8/1k4pp/3K4/5p2/7P/5PP1/8/8 w - - 0 1	?	g4′	?	**g4′′**	DTZ decides without EGT: immediate P-push ⇒ *dtz* = 1p
09	1.12	24b	8/5k2/3Q4/pp4K1/1P6/P7/8/8 b - - 0 1	Kg7′	Kg7′	Kg7′	**Kg7′′**	DTC/Z decide without EGT: axb4 and a4 are both rejected
10	1.14	15b	7Q/pP6/Pk6/1P6/6K1/8/8/8 b - - 0 1	Kc5′	**Kc5′′**	—	—	*dtm* < *dtf*: *dtm* = 8p and *dtf* = 16p
11	1.14	16w	7Q/pP6/P7/1Pk5/6K1/8/8/8 w - - 0 1	b8Q′	Qb2′	**b8Q′′**	—	FINALGEN cannot contemplate endgame KQQPPKP
12	2.03	25w	8/2P5/pK1k4/p7/8/1P6/8/8 w - - 0 1	**c8Q′′**	—	—	—	… and not CRAFTY's 25. Kxa5, reaching for the 5m EGTs
13	1.12	20w	6k1/8/p4P1P/1p4K1/1P6/P7/8/8 w - - 0 1	?	**f7′′**	—	—	Unnatural: 20. Kg6 promotes a pawn quickly
14	1.12	20b	6k1/5P2/p6P/1p4K1/1P6/P7/8/8 b - - 0 1	?	**Kf8′′**	—	—	Unnatural: 20. … Kxf7 clearly prolongs the line
15	1.09	3b	8/1k4pp/3K4/8/6PP/8/8/8 b - - 0 1	Kb6′	Kb6′	?	**Kb6′′**	DTF-excluded, DTZ-u-optimal Kc8! requires Ke(6/7) ′′′′
16	1.15	17b	1Q6/2k5/4P3/5P2/7p/6p1/7P/K7 b - - 0 1	?	**Kc6′′**	—	—	Kc6 '⇒' *dtf* = 24p though *dtm* = ~12p. Kxb8 is more natural

Table 3. Some illustrative positions and decisions taken by engine *E*.

Considerations of the two key resources, space and time, suggest that the EH/HM_5 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 *E* inherits the unlimited space/time resources used to compute EGTs whereas HM_5 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 CRAFTY would have to progress before invoking a '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 *v* of 'the move' can be assessed by analyzing the TS1/2 positions with the other side to move (ostm). Verdicts after '⯑' are not purely EGT-based but required some tree-search and evaluation:

- *v* = 1 point, win becomes a loss: TS2.01; ⯑ TS1.12/15, TS2.06/13
- *v* = ½ point, win becomes a draw: TS1.01/05-11, TS2.03/08-11; ⯑ TS1.14, TS2.04-05/15-16
- *v* = ½ point, draw becomes a loss: TS2.02
- *v* = 0 point, result unchanged:
  - TS1.02-04, TS2.12. Note that TS1.03/04 and TS2.12 are type-BM zugs;[6]
  - ⯑ TS1.13/16, TS2.07/14

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 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 FINALGEN and chess-engine search, so the exclusive use of KP-endgames 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 HHDBIV (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?

[6] A type-BM zug is one in which DTM is greater with the move than without it (Bleicher and Haworth, 2010)

#	id	HHdbIV #	HHdbIV pos.	Author(s)	Year	m	w-b	GBR code	s8m	F	Position: 'FEN' Forsyth Extended Notation	Val.	DEEP FRITZ 14, 2-core, 3m Eval.	Δ	Depth	Choice	✓/✗
01	H01	4728	1w	Behting	1906	9	4-5	0002.14	n	n	8/8/7p/3KNN1k/2p4p/8/3P2p1/8 w - - 0 1	=	-4.10	1.12	21	1. Ng7+	✗
02	H02	25700	6w	Gurvich	1952	7	4-3	0044.10	Y	n	1N6/6k1/1B6/8/1P6/8/b1K5/n8 w - - 0 1	1-0	4.35	0.44	25	6. Kd1	✗
03	H03	70286	3w	Antonini	2003	7	3-4	4001.02	Y	n	8/8/2K5/4N3/p5Q1/4k2p/8/q7 w - - 0 1	1-0	12.20	0.00	19	3. Qf3+	✓
04	H04	56516	4w	Pervakov & Grin	1988	7	3-4	1400.02	Y	n	8/8/8/5R2/8/kr1p2p1/3K3Q w - - 0 1	1-0	∞	∞	21	4. Qh2	✓
05	H05	56539	4w	Kuryatnikov	1988	7	4-3	0041.11	Y	n	1k6/2p4B/4b3/4N3/2K5/8/P7/8 w - - 0 1	1-0	7.41	6.80	27	4. Kb5	✓
06	H06	35997	6w	Mitrofanov	1967	9	5-4	4033.30	n	n	kb4Q1/P7/1PP5/K6q/8/8/8/4n3 w - - 0 1	1-0	12.65	12.65	21	6. Qg5	✓
07	H07	8860	6w	Zepler	1923	7	4-3	0400.21	Y	n	3k4/3P4/P2R4/8/6p1/r7/4K3/8 w - - 0 1	1-0	3.06	0.00	29	6. Ke1	✓
08	H08	67602	1w	Smyslov	2000	8	4-4	0000.33	n	Y	8/8/6p1/5p2/p1k2P2/8/P6P/4K3 w - - 0 1	=	-0.66	11.04	26	1. a3	✓
09	H09	39037	1w	Bazlov	1971	7	4-3	0405.00	Y	n	n7/2N4K/8/k7/7R/2r5/8/N7 w - - 0 1	1-0	6.88	6.02	20	1. Rh5+	✓
10	H10	67600	2w	Smyslov	2000	9	4-5	0130.23	n	Y	8/5k1p/4p3/2K4P/5P2/3b4/p7/6R1 w - - 0 1	=	0.00	0.62	22	2. h6	✓
11	H11	32098	1w	Fritz	1961	7	4-3	0310.21	Y	Y	B7/P7/P7/k6K/8/8/7p/r7 w - - 0 1	1-0	3.40	10.93	29	1. Bh1	✓
12	H12	67945	2w	v. d. Heijden & Beasley	2000	7	5-2	0013.30	Y	Y	8/8/8/8/8/K1P5/PBPn4/1k6 w - - 0 1	1-0	6.98	6.62	21	2. c4	✓
13	H13	57384	1w	Arestov	1989	12	6-6	0053.33	n	n	2K5/4p1B1/4k1P1/1b3pP1/p3n3/3P4/4B3/8 w - - 0 1	1-0	1.21	0.46	21	1. Bb2	✗
14	H14	58049	5w	Neishtadt	1989	16	8-8	3213.45	n	n	7k/p4p1B/p4P2/P3qP2/7R/p1p2R2/P7/Kn6 w - - 0 1	=	0.00	4.93	22	1. Re3	✓
15	H15	69180	6w	Fabiano	2002	12	6-6	3001.44	n	Y	3N4/K3p3/4p3/k1P5/p1P3p1/P7/4P3/7q w - - 0 1	1-0	10.96	10.97	19	1. e3	✓
16	H16	69110	4w	Kralin	2002	13	6-7	0130.45	n	Y	5bRK/6p1/2p4k/2P1p1p1/4p1P1/4P3/2P5/8 w - - 0 1	1-0	0.00	0.00	34	1. c3	✓
17	V01	32412	1w	Zaitsev	1962	8	4-4	0107.11	n	n	5Kn1/4n3/8/5P2/8/3k4/3p4/1N3R2 w - - 0 1	1-0	0.69	0.21	24	1. Rd1	✗
18	V02	71075	1w	Gurgenidze & Kalandadze	2004	12	7-5	0300.63	n	Y	8/pPPp1p2/3P1PPr/8/2P5/2k5/8/2K5 w - - 0 1	1-0	16.04	16.04	23	1. Kd1	✓
19	V03	64369	1w	Gurgenidze & Kalandadze	1997	13	7-6	0801.33	n	n	4R3/k6r/P3p3/K7/5P2/P5pp/8/RN1r4 w - - 0 1	=	0.00	10.40	18	1. Nc3	✓
20	V04	55531	1w	Gurgenidze	1987	12	7-5	0613.51	n	n	5Bk1/1PP5/5P2/7P/n6p/4r2P/r7/6K1 w - - 0 1	1-0	6.09	∞	24	1. Bb4	✓
21	V05	—	1w	Salai	2011	12	7-5	0040.53	n	Y	4K1k1/8/1p5p/1Pp3b1/8/1P3P2/P1B2P2/8 w - - 0 1	1-0	1.24	0.30	26	1. Kd7	✗
22	V06	71074	1w	Benno	2004	10	7-3	3411.30	n	n	k1N5/1r1P4/8/KP6/7q/P7/6RB/8 w - - 0 1	1-0	15.79	0.00	22	1. Rg8	✓
23	V07	69009	1w	Visokosov	2002	14	7-7	0071.44	n	n	B2k4/3Pp3/4P1P1/5p1P/1pK5/7b/5p2/2b2N2 w - - 0 1	1-0	0.28	0.28	23	1. g7	✗
24	V08	72886	1w	Kovalenko	2006	13	6-7	0000.56	n	Y	8/6Pp/p4K1p/P6p/8/P6p/Ppk4P/8 w - - 0 1	1-0	6.38	5.98	17	1. g8Q	✓
25	V09	72995	1w	Sochniev	2006	8	3-5	0034.12	n	n	2n5/1p2p2N/6P1/8/7k/2K5/8/6b1 w - - 0 1	1-0	2.42	3.59	23	1. g7	✓
26	V10	72386	1w	Csengeri	2005	11	6-5	0700.42	n	n	7K/1p1R4/1pP5/8/P7/P7/P1k3r1/3r4 w - - 0 1	1-0	3.29	4.08	23	1. c7	✓
27	V11	57418	1w	Pervakov & Sumbatyan	1989	22	12-10	3812.66	n	n	1N6/pq3P2/p2p4/P3k1B1/2RNp1r1/P1P1P3/2P1p1rp/4R2K w - - 0 1	1-0	9.11	9.11	17	1. Bf4+	✓
28	V12	38172	1w	Alekseev	1970	18	10-8	4825.23	n	n	1K1n3B/r6q/pR6/k1pN1R2/p1P4r/P1N4B/Q7/8 w - - 0 1	1-0	12.66	0.32	19	1. Qb1	✓
29	V13	66438	1w	Fiedler	1999	21	11-10	3111.78	n	n	8/1p2Pq1B/3p2p1/3N1kp1/1P1P4/2pP2P1/p1Pp1pP1/5R1K w - - 0 1	1-0	3.69	13.55	22	1. Bg8	✓
30	V14	73873	1w	Katsnelson & Sochniev	2007	8	3-5	0113.03	n	n	8/K7/1B1n4/8/1p5k/p4p2/6R1/8 w - - 0 1	=	-3.43	4.09	18	1. Bf2+	✓
31	V15	73067	1w	Ryabinin	2006	12	7-5	0300.63	n	Y	8/2P3P1/1pPp4/p7/4P1PP/7K/4k3/5r2 w - - 0 1	1-0	9.28	6.96	21	1. g5	✓
32	V16	75276	1w	Didukh & Masimov	2009	12	6-6	0311.34	n	n	8/5p2/5P2/P7/3B4/8/pp2KPp1/r1k1N3 w - - 0 1	1-0	0.00	18.84	17	1. Be3+	✓
33	V17	61165	1w	Pervakov & Selivanov	1993	12	5-7	4070.23	n	n	b2q4/1kbP2p1/1B6/1QP5/8/7p/p7/7K w - - 0 1	1-0	5.49	∞	16	1. c6+	✓
34	V18	59763	1w	Elkies	1991	14	8-6	3002.54	n	n	8/1p6/1p6/kPp2P1K/2P5/N1Pp4/q2P4/1N6 w - - 0 1	=	-0.17	0.00	23	1. f6	✓
35	H17	3477	5b	Saavedra & Barbier	1895	4	2-2	0300.10	Y	Y	8/2P5/8/8/8/3r4/2K5/k7 b - - 0 1	1-0	∞	0.00	28	5... Rf3	✗
36	H18	69110	1w	Kralin	2002	16	8-8	4130.55	n	n	4R2K/3qb1p1/2p3kP/2P1p1p1/4p3/4P2P/2P5/6Q1 w - - 0 1	1-0	0.00	∞	24	1. Rg8	✓
37	H19	—	1w	Krug	2012	18	8-10	4062.46	n	n	Q3b3/1p2q3/1p6/bp5P/1p6/1p6/1P1N1PPp/4NK1k w - - 0 1	1-0	0.00	0.00	21	1. f3	✗

Table 4. The studies cited by van der Heijden (2006, 2014) and Vlasák (2013).

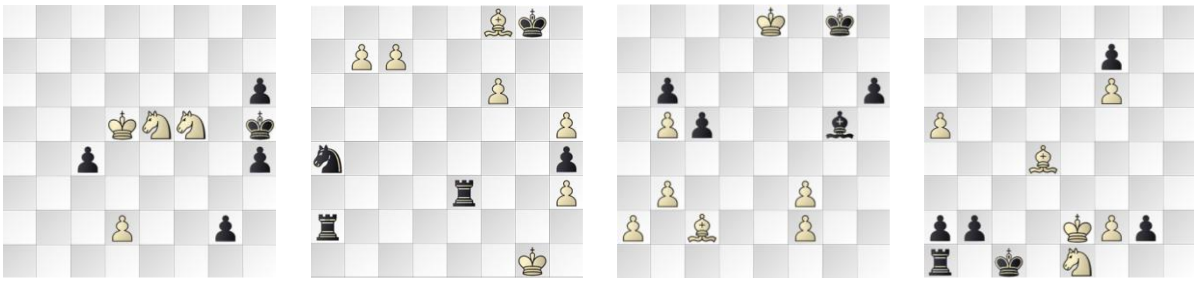

Figure 3. Four studies from Table 4: H01 (White to draw); V04, V05 and V16 (White to win).

Table 4 lists the 37 positions that van der Heijden and Vlasák chose. Authors are credited and serial numbers in the HHDBIV corpus are given.[7] 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 'Δ' 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-test studies with a battery of significantly different engines. Vlasák used HIARCS and HOUDINI, HIARCS 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 presumes a fallible opponent who can be tricked

[7] HHDBIV indices for the 16-study subset of TS1/2 are: TS1.01-04 (#7988, 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.[8] 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[9] 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

Bleicher, E. and Haworth, G. M^c^C. (2010) 6-man chess and zugzwangs. Advances in Computer Games 12, Pamplona. LNCS 6048. pp. 123-135. Springer. ISSN 0302-9743 doi: 10.1007/978-3-642-12993-3_12.

CPW (2013). https://chessprogramming.wikispaces.com/Ronald+de+Man. Ronald de Man.

CPW (2015). https://chessprogramming.wikispaces.com/Gyula+Horváth. Gyula Horváth.

Fine, R. (1941). *Basic Chess Endings*. David McKay. Revised edition, ISBN 978-0812934939 (2003).

Haworth, G. McC. (2013a). Haworth's Law. *ICGA Journal*, Vol. 36, No. 4, p. 230.

Haworth, G. McC. (2013b). Chess Endgame News. *ICGA Journal*, Vol. 36, No. 4, pp. 228-9.

Haworth, G. McC. (2015). This note plus supporting datasets. http://centaur.reading.ac.uk/39946/.

Hyatt, R. (2015). https://cis.uab.edu/hyatt/crafty/. Hyatt's CRAFTY website.

MVL (2015). http://tb7.chessok.com/probe/250/3232. Lomonosov server of s8m DTM data.

Nalimov, E. V., Haworth, G. McC., and Heinz, E. A. (2000). Space-efficient Indexing of Chess Endgame Tables. *ICGA Journal*, Vol. 23, No. 3, pp. 148-162.

Newborn and Hyatt (2014). Computer Chess Endgame Play with Pawns: Then and Now. *ICGA Journal*, Vol. 37, No. 4, pp. 195-206.

Nunn, J. (2012). The Behting study is sound. Chessbase's Chess News: http://preview.tinyurl.com/ns57pkl.

Romero, P. P. (2012). FINALGEN: download, tutorial and examples. http://www.mtu-media.com/finalgen.

Tamplin, J. (2015). http://chess.jaet.org/endings/. Multi-metric endgame table service.

Van der Heijden, H. M. J. F. (2006). Schaaksoftware: eindspelstudietest. *EBUR* Vol. 18, No. 4, pp. 19-21.

Van der Heijden, H. M. J. F. (2010). http://www.hhdbiv.nl/. HHDBIV, ENDGAME STUDY DATABASE IV.

Van der Heijden, H.M.J.F. (2014). Onoplosbare (?) eindspelstudies [Insoluble (?) endgame studies]. *Computerschaak*. Vol. 33, No. 1, pp. 10-11.

Vlasák, E. (2013) Computer News: Looking far ahead. *EG*, Vol. 19, No. 193, pp. 219-226.

[8] This is a reflection on the concept of 'correctness'. For *H17*, engines choose **5. … Rf3** postponing mate, but the swindle **5. … Rd4** hopes for **6. 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. Rff7**. HHDBIV notes these flaws where known and includes repaired versions of studies where available.