FACTORS AFFECTING DIMINISHING RETURNS FOR SEARCHING DEEPER[1]

Matej Guid[2] *and Ivan Bratko*[2]

Ljubljana, Slovenia

ABSTRACT

The phenomenon of diminishing returns for additional search effort has been observed by several researchers. We study experimentally additional factors which influence the behaviour of diminishing returns that manifest themselves in *go-deep* experiments. The results obtained on a large set of more than 40,000 positions from chess grandmaster games using the programs CRAFTY, RYBKA, and SHREDDER show that diminishing returns depend on (a) the values of the positions, (b) the quality of the evaluation function of the program used, and to some extent also on (c) the phase of the game, and the amount of material on the board.

1 INTRODUCTION

Deep-search behaviour and diminishing returns for additional search in chess have been burning issues in the last twenty five years in the game-playing scientific community. Two different approaches took place on this topic: *self-play* and *go-deep*. While in self-play experiments, two otherwise identical programs are matched with one having a handicap (usually in search depth), go-deep experiments deal with best-move changes resulting from different search depths of a set of positions.

The go-deep experiments were introduced for determining the expectation of a new best move being discovered by searching only one ply deeper. The approach is based on Newborn's (1985) discovery that the results of self-play experiments are closely correlated with the rate at which the best move changes from one iteration to the next. Newborn (1985) formulated the following hypothesis. Let $RI(d + 1)$ denote the rating improvement when increasing the search depth from level d to level $d + 1$, and $BC(d)$ the expectation of finding a best move at level d different from the best move found at level $d - 1$, then:

$$RI(d+1) = \frac{BC(d+1)}{BC(d)} \cdot RI(d) \qquad (1)$$

There were some objections about the above equation, e.g., the one by Heinz (1998): "Please imagine a chess program that simply switches back and forth between a few good moves all the time. Such behaviour does surely not increase the playing strength of the program at any search depth." He suggested that the discovery of "fresh ideas" looks like a much better and meaningful indicator of increases in playing strength than a best-move change at the next iteration of the search, and proposed "fresh best" moves instead, defined as new best moves which the program never deemed best before. Whatever the merit of this proposal, determining $BC(d)$ for higher values of d continued to be used in several experiments. In 1997, PHOENIX (Schaeffer, 1988) and THE TURK (Junghanns *et al*,. 1997) were used to record best-move changes at iteration depths up to 9 plies. In the same year, Hyatt and Newborn (1997) let CRAFTY search to an iteration depth of 14 plies. In 1998, Heinz (1998) repeated their go-deep experiment with DARKTHOUGHT. All these experiments were performed on somehow limited

[1] This article is a revised version of a contribution with the same title to the Computer Games Workshop 2007 held in Amsterdam, The Netherlands, June 15-17, 2007. The revised version has been subjected to the Journal's normal refereeing procedure.

[2] Artificial Intelligence Laboratory, Faculty of Computer and Information Science, University of Ljubljana, Slovenia. Email: {matej.guid,bratko}@fri.uni-lj.si.

datasets of test positions and did *not* provide any conclusive empirical evidence that the best move changes taper off continuously with increasing search depth.

An interesting go-deep experiment was performed by Sadikov and Bratko (2006). They made very deep searches (unlimited for all practical purposes) possible by concentrating on chess endgames with a limited number of pieces. Their results confirmed that diminishing returns in chess exist, and showed that the amount of knowledge, which a program has, influences the precise time when diminishing returns will start to manifest themselves.

A remarkable follow-up on the previous work done on deep-search behaviour using chess programs was published by Steenhuisen (2005) who used CRAFTY to repeat the go-deep experiment on positions taken from previous experiments to push the search horizon to 20 plies. He used the same experimental setup to search, among others, a set of 4,500 positions, from the opening phase, to a depth of 18 plies. His results show that the chance of new best moves being discovered decreases exponentially when searching to higher depths, and decreases faster for positions closer to the end of the game. He also reported that the speed with which the best-change rate decreases depends on the test set used.

The latter seems to be an important issue regarding the trustworthiness of the various results obtained by the go-deep experiments. How can one rely on statistical evidence from different go-deep experiments, if they obviously depend on the dataset used? In this article we address that issue, and investigate the hypothesis that the rate at which the returns diminish depends on the value of the position. Using a large dataset of more than 40,000 positions taken from real games we conduct go-deep experiments with the programs CRAFTY, RYBKA, and SHREDDER to provide evidence that the chance of new best moves being discovered at higher depths depends on:

1. the values of positions in the dataset,
2. the quality of the evaluation function of the program used,

and to some extent also on

3. the phase of the game, and the amount of material on the board.

2 GO-DEEP EXPERIMENT

The Chess programs CRAFTY, RYBKA, and SHREDDER[3] were used to analyse more than 40,000 positions from real games played in World Chess Championship matches (the WCC dataset). Each position occurring in these games after move 12 was searched to a fixed depth ranging from 2 to 12 plies[4].

For the measurements done in the go-deep experiments we use the same definitions as provided by Heinz (1998) and Steenhuisen (2005). Let $B(d)$ denote the best move after a search to depth d, then the following best-move properties were defined.

Best Change $B(d) \neq B(d - 1)$
Fresh Best $B(d) \neq B(j) \ \forall j < d$
(d-2) Best $B(d) = B(d - 2)$ and $B(d) \neq B(d - 1)$
(d-3) Best $B(d) = B(d - 3)$ and $B(d) \neq B(d - 2)$ and $B(d) \neq B(d - 1)$

We give the estimated probabilities (in %) and their estimated standard errors SE (in Equation 2: $N(d)$ stands for the number of observations at search depth d) for each measurement of Best Change. The rates for Fresh Best, (d - 2) Best, and (d - 3) Best are given as conditional to the occurrence of a Best Change. We also provide mean evaluations of positions at each level of search.

[3] CRAFTY 19.2, RYBKA 2.2n2, and DEEP SHREDDER 10 UCI were used in the experiments.
[4] More details about the chosen experimental setup could be found in Guid and Bratko (2006). Search to depth 1 was omitted due to specific limitations of the programs used. The complete set of games used for the analysis could be found on first author's website: http://www.ailab.si/matej/.

$$SE = \sqrt{\frac{BC(d)(1-BC(d))}{N(d)-1}} \tag{2}$$

For confidence bounds on the values for best-change rates we use the 95%-level of confidence (λ = 1.96). Moreover, we use the equation given by Steenhuisen (2005) (in Equation 3: *m* represents the number of successes in a sample size of *n* observations).

$$\frac{m+\frac{\lambda^2}{2} \pm \sqrt{m(1-\frac{m}{n})+\frac{\lambda^2}{4}}}{n+\lambda^2} \tag{3}$$

Our hypothesis is the following: best-move changes depend on the value of a given position. It was based on an observation that move changes tend to occur more frequently in balanced positions. To determine the best available approximation of "the true value" of each analysed position, the evaluation at depth 12 served as an oracle. We devised different groups of positions based on their estimated true values.

The rest of the paper is organised as follows. Sections 3 and 4 present the results of go-deep experiments performed by CRAFTY and RYBKA on different groups of positions, based on their estimated true values. Section 5 gives a comparison of Best-Change rates of the programs. In Sections 6 and 7 we observe best-move changes in balanced positions of different groups, based on the phase of the game and the number of pieces on the board. Properties of the groups of positions are described at the beginning of each of these sections. We summarise our results in Section 8.

3 CRAFTY GOES DEEP

Several researchers have used CRAFTY for their go-deep experiments. However, none had such a large set of test positions at his/her disposal as we have (over 40,000 positions). Steenhuisen (2005) observed deep-search behaviour of CRAFTY on different test sets and reported different best-change rates and best-change rate decreases for different test sets. This and the following section will show that best-change rates strongly depend on the values of the positions included in a test set.

We divided the original test set into six subsets, based on the evaluations of the positions obtained at depth 12 as presented in Table 1. In the usual terms of chess players, the positions of Groups 1 and 6 could be labelled as positions with a "decisive advantage", those of Groups 2 and 5 with a "large advantage", while Groups 3 and 4 consist of positions regarded as approximately equal or with a "small advantage" at most.

Group	1	2	3	4	5	6
Evaluation(x)	$x<-2$	$-2\leq x<-1$	$-1\leq x<0$	$0\leq x<1$	$1\leq x<2$	$x\geq 2$
Positions	4,011	3,571	10,169	18,038	6,008	6,203

Table 1: Subsets with positions of different range of evaluations obtained at level 12 (CRAFTY).

The results for each of the six groups are presented in Figure 1. The curves clearly show a different deep-search behaviour of the program for the different groups, depending on the estimated value of positions they consist of. The chance of new best moves being discovered at higher depths is significantly higher for balanced positions than for positions with a decisive advantage. It is interesting to observe that this phenomenon does not yet occur at the shallowest search depths, while in the results of RYBKA it manifests itself at each level of search (see Section 4).

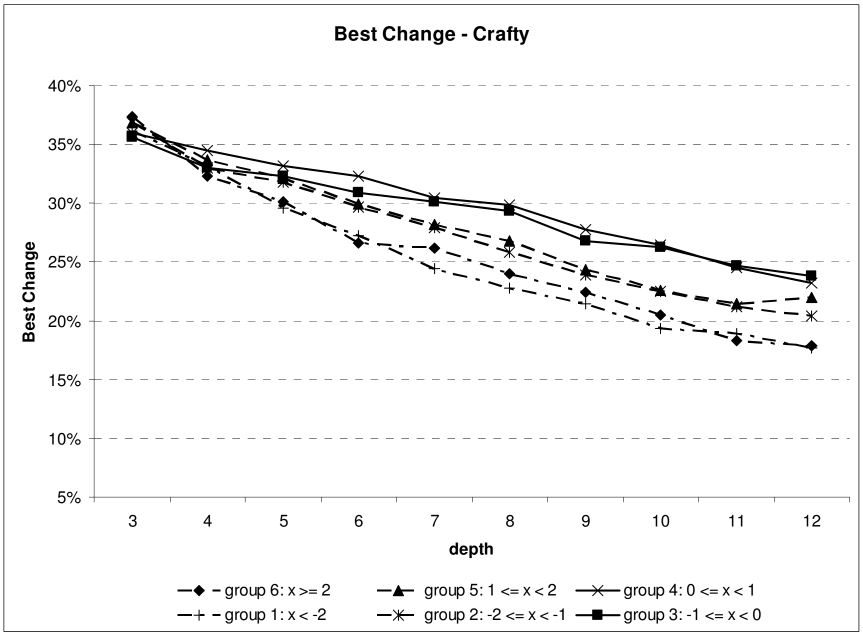

Figure 1: Go-deep results of CRAFTY on the six different groups of positions.

Tables 2 and 3 show the best-move properties for Groups 4 and 6. While the results resemble the ones obtained by Steenhuisen (2005) on the 4,500 positions of the ECO test set in a sense that both Best-Change and Fresh-Best rates decrease consistently with increasing search depth, the rates nevertheless significantly differ for each of the two groups of positions.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	35.96 (0.36)	100.00	-	-	0.36
4	34.47 (0.35)	74.88	25.12	-	0.37
5	33.18 (0.35)	64.16	27.34	8.50	0.37
6	32.34 (0.35)	54.38	28.44	11.38	0.37
7	30.48 (0.34)	49.53	31.14	9.51	0.37
8	29.86 (0.34)	42.81	31.45	11.27	0.38
9	27.75 (0.33)	40.02	33.87	10.81	0.38
10	26.48 (0.33)	37.77	33.31	10.57	0.38
11	24.53 (0.32)	34.79	33.48	11.14	0.38
12	23.17 (0.31)	32.26	33.07	12.04	0.39

Table 2: Results of CRAFTY for the 18,038 positions of Group 4.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	37.42 (0.61)	100.00	-	-	2.64
4	32.27 (0.59)	73.93	26.07	-	2.76
5	30.13 (0.58)	64.85	24.83	10.33	2.84
6	26.60 (0.56)	55.70	28.06	9.70	2.95
7	26.21 (0.56)	49.88	27.37	10.52	3.04
8	23.99 (0.54)	39.92	31.18	11.02	3.17
9	22.44 (0.53)	37.21	32.18	12.72	3.29
10	20.47 (0.51)	36.30	30.79	11.50	3.42
11	18.30 (0.49)	31.37	32.42	12.07	3.54
12	17.85 (0.49)	29.27	29.99	13.91	3.68

Table 3: Results of CRAFTY for the 6,203 positions of Group 6.

The 95%-confidence bounds for Best Change (calculated using the Equation 2 given in Section 2) at the highest level of search performed for the samples of 18,038 and 6,203 positions of Groups 4 and 6 are [22.56;23.97] and [16.91;18.82], respectively.

4 RYBKA GOES DEEP

RYBKA is currently the strongest chess program given its WCCC 2007 title in Amsterdam (see elsewhere in this issue) and according to the SSDF rating list (see p. 127 of this issue). To the best of our knowledge there were no previous go-deep experiments performed with this program. The results in this section will not only confirm that best-change rates depend on the values of the positions, but also demonstrate that the chance of new best moves being discovered at higher depths is lower at all depths compared to CRAFTY, which is rated more than 300 rating points lower on the aforementioned rating list. Table 4 presents the subsets evaluated by RYBKA, analogous to those presented in Table 1 and evaluated by CRAFTY.

Group	1	2	3	4	5	6
Evaluation(x)	x<-2	-2≤x<-1	-1≤x<0	0≤x<1	1≤x<2	x≥2
Positions	1,263	1,469	9,808	22,644	3,152	2,133

Table 4: Subsets with positions of different range of evaluations obtained at level 12 (RYBKA).

The results of RYBKA presented in Figure 2 resemble the results of CRAFTY in Figure 1, except that all the curves appear significantly lower on the vertical scale. This result seems to be in line with the observation, based on the results by Sadikov and Bratko (2006), that the amount of knowledge a program has (or the quality of the evaluation function) influences the deep-search behaviour of a program. The big difference in strength of the two programs is likely to be the consequence of RYBKA having a stronger evaluation function; it is as well commonly known that chess players prefer evaluations of this program to CRAFTY's evaluations. In their study, Sadikov and Bratko (2006) claim that diminishing returns will start to manifest themselves earlier using a program with a stronger evaluation function, based on experiments performed on chess endgames, at the same time suspecting that similar results would be obtained with more pieces on the board. The results presented here seem to be in accordance with that conjecture.

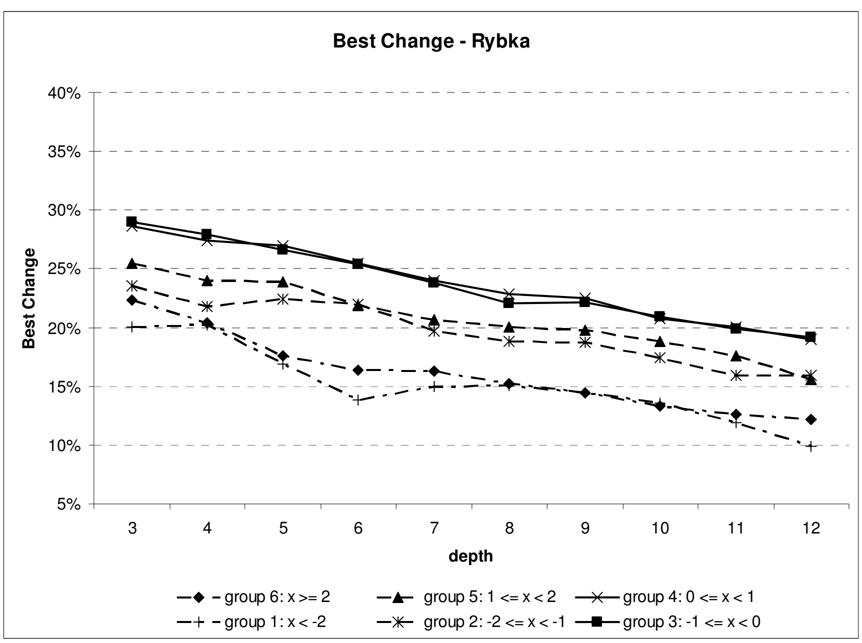

Figure 2: Go-deep results of RYBKA on the six different groups of positions.

It is also interesting to observe that the mean evaluations of both programs in won positions monotonically increase with increasing search depth. Similarly, we observed that in lost positions (where negative evaluations are used), the mean evaluations monotonically decrease by searching more deeply. We believe that this phenomenon represents a direct consequence of a desired property of heuristic evaluation functions: to reflect a progress towards the game-theoretical result of the game. Tables 5 and 6 (results of RYBKA) are the analogons of Tables 2 and 3 (results of CRAFTY).

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	28.59 (0.30)	100.00	-	-	0.31
4	27.36 (0.30)	71.42	28.58	-	0.31
5	27.00 (0.30)	62.95	27.12	9.93	0.31
6	25.44 (0.29)	53.32	28.13	10.45	0.31
7	24.00 (0.28)	49.91	26.63	11.21	0.30
8	22.88 (0.28)	45.78	26.85	11.37	0.30
9	22.50 (0.28)	42.97	25.63	11.46	0.30
10	20.73 (0.27)	37.17	28.46	11.31	0.30
11	20.03 (0.27)	36.16	27.76	11.78	0.30
12	19.01 (0.26)	34.08	27.87	11.85	0.30

Table 5: Results of RYBKA for the 22,644 positions of Group 4.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	22.36 (0.90)	100.00	-	-	2.49
4	20.39 (0.87)	77.24	22.76	-	2.60
5	17.63 (0.83)	66.76	24.20	9.04	2.77
6	16.41 (0.80)	54.86	25.43	10.57	2.89
7	16.32 (0.80)	49.71	26.44	10.06	3.01
8	15.24 (0.78)	44.00	23.69	13.23	3.14
9	14.49 (0.76)	45.63	24.60	10.36	3.27
10	13.31 (0.74)	42.61	23.94	12.68	3.42
11	12.61 (0.72)	37.92	24.16	8.55	3.59
12	12.19 (0.71)	36.54	30.00	7.31	3.75

Table 6: Results of RYBKA for the 2,133 positions of Group 6.

The 95%-confidence bounds for Best Change at the highest level of search performed for the samples of 22,644 and 2,133 positions of Groups 4 and 6 are [18.51;19.53] and [10.87;13.65], respectively.

5 DIMINISHING RETURNS AND QUALITY OF EVALUATION FUNCTION

In this section, we give a comparison of the Best-Change rates of the programs. In order to verify our hypothesis that best-move changes correlate with the quality of the evaluation function of a program, we used another program, SHREDDER, to analyse more than 40,000 positions from the WCC dataset, using the same methodology. SHREDDER is currently one of the strongest chess programs, however, it is commonly accepted among strong chess players that its evaluations are somehow less reliable than those of RYBKA[5]. We also did "SHREDDER goes deep" experiments, and the results show the same trends as those obtained by CRAFTY and RYBKA.

Figure 3 shows Best-Change rates of the three programs on approximately equal positions of Group 4. Qualitatively, similar results were observed for other groups of positions as well. RYBKA (which is regarded as the program with the strongest evaluation function of the three programs) has the lowest Best-Change curves, and the opposite applies to CRAFTY (whose evaluation function is considered to be the weakest one).

[5] The April 2007 SSDF rating list (http://ssdf.bosjo.net/list.htm) (with slightly different versions of RYBKA and CRAFTY) gives the following ratings: RYBKA 2962, SHREDDER 2830, and CRAFTY 2614.

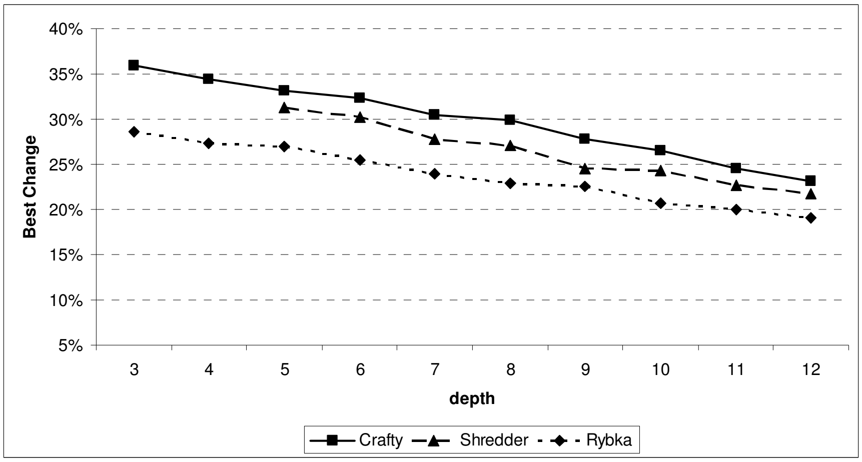

Figure 3: Go-deep results of CRAFTY, SHREDDER[6], and RYBKA on (approximately equal) positions of Group 4.

Could Best-Change rates be used as a direct measure of the quality of an evaluation function? This question requires a further investigation. Consider a program that always selects the first move of the alphabetically sorted possible moves. In such a case, the Best-Change curve coincides with the horizontal axis (which also happens in case of a perfect evaluation), despite of terribly low quality of such program's evaluations. However, this is a rather contrived case. As our experimental results suggest, evaluation functions of successful programs may have some properties that make comparison based on Best-Change rates sensible. In each case, we suggest that a dataset used for such an investigation should be representative for the whole game of chess, as it was probably the case with our large dataset of real-game positions.

6 DIMINISHING RETURNS AND PHASE OF THE GAME

Steenhuisen (2005) was the first to point out that the chance of new best moves being discovered at higher depth decreases faster for positions closer to the end of the game. However, having in mind that deep-search behaviour depends on the values of positions in a test set, it seems worthwhile to check whether his results were just the consequence of dealing with positions with a decisive advantage (at least on average) in a later phase of the game. For the purpose of this experiment we took only a subset with more or less balanced positions with depth 12 and an evaluation in the range between -0.50 and 0.50 (see Table 7). Our results show that in the positions that occurred in the games later than move 50, the chance of new best moves being discovered indeed decreases faster, which agrees with Steenhuisen's (2005) observations. The experiments in this and the following section were performed by CRAFTY.

Group	1	2	3	4	5
Move no.(x)	x<20	20≤x<30	30≤x<40	40≤x<50	x≥50
Positions	7,580	6,106	3,418	1,356	961

Table 7: Five subsets of positions of different phases in the game, with evaluations in range between -0.50 and 0.50, obtained at search depth 12.

The results presented in Figure 4 show that while there is no obvious correlation between move number and the chance of new best moves being discovered at higher depth, in the positions of Group 5 that occurred closer to the end of the game the Best-Change curve nevertheless appears lower than the curves of the other groups. Tables 8 and 9 show the best-move properties for the Groups 1 and 5.

[6] The version of the program we used provides evaluations only from depth 4 onwards.

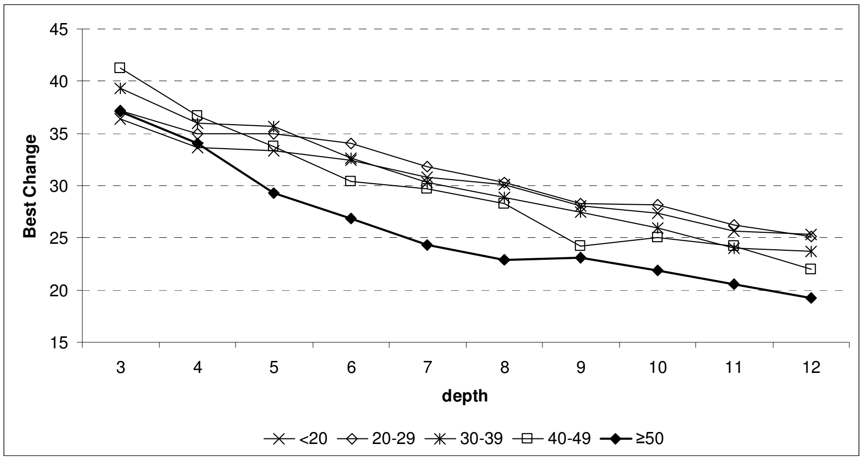

Figure 4: Go-deep results with positions of different phases of the game.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	36.41 (0.55)	100.00	-	-	0.08
4	33.63 (0.54)	75.56	24.44	-	0.08
5	33.35 (0.54)	63.57	27.73	8.70	0.08
6	32.39 (0.54)	54.01	29.74	10.55	0.08
7	30.83 (0.53)	49.64	31.79	9.33	0.07
8	30.08 (0.53)	44.65	31.49	10.22	0.07
9	28.10 (0.52)	41.60	31.64	10.47	0.07
10	27.40 (0.51)	37.31	34.28	10.01	0.07
11	25.66 (0.50)	35.73	34.91	10.03	0.07
12	25.32 (0.50)	31.16	34.55	11.52	0.07

Table 8: Results for the 7,580 positions of Group 1.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	37.04 (1.56)	100.00	-	-	0.07
4	34.03 (1.53)	72.78	27.22	-	0.05
5	29.24 (1.47)	60.85	27.40	11.74	0.05
6	26.85 (1.43)	49.22	30.23	14.34	0.03
7	24.35 (1.39)	47.44	29.91	10.26	0.02
8	22.89 (1.36)	45.91	27.27	9.55	0.02
9	23.10 (1.36)	38.29	32.88	10.81	0.02
10	21.85 (1.33)	37.62	27.62	11.43	0.02
11	20.60 (1.31)	33.33	32.83	12.12	0.02
12	19.25 (1.27)	26.49	36.22	8.65	0.01

Table 9: Results for the 961 positions of Group 5.

The 95%-confidence bounds for Best Change at the highest level of search performed for the samples of 7,580 and 961 positions of Groups 1 and 5 are [24.35;26.31] and [16.88;21.86], respectively.

7 DIMINISHING RETURNS AND MATERIAL

The phase of the game is closely correlated with the amount of material on the board. Therefore, in accordance with previous observations, it could be expected that the rate of best-change properties will be lower in positions with fewer pieces on the board. The results of this section confirm that with a total value of pieces less than 15 for each of the players, the chance of new best moves being discovered at higher depth decreases faster, albeit only from depth 5 on (also the differences are not so obvious as in the previous section). In the total value of the pieces, the Pawns are counted in and for the values of pieces the commonly accepted values are taken (Queen = 9, Rook = 5, Bishop = 3, Knight = 3, Pawn =

1). Table 10 shows a division into six subsets, determined by the amount of material present at the board.

Group	1	2	3	4	5	6
Material(x)	x<15	15≤x<20	20≤x<25	25≤x<30	30≤x<35	x≥35
Positions	3,236	1,737	2,322	2,612	4,082	4,112

Table 10: Six subsets of positions with different amount of material on the board (each player starts with the amount of 39 points), with evaluations in range between -0.50 and 0.50, obtained at search depth 12.

Figure 5 shows that material and best-move changes are not clearly correlated. It is only the curve for positions with the total piece value of less than 15 points of material (for each of the players) that slightly deviates from the others. Surprisingly, we did not spot any significant deviations in positions with even less material either. Tables 11 and 12 show the best-move properties for Group 6 and Group 1.

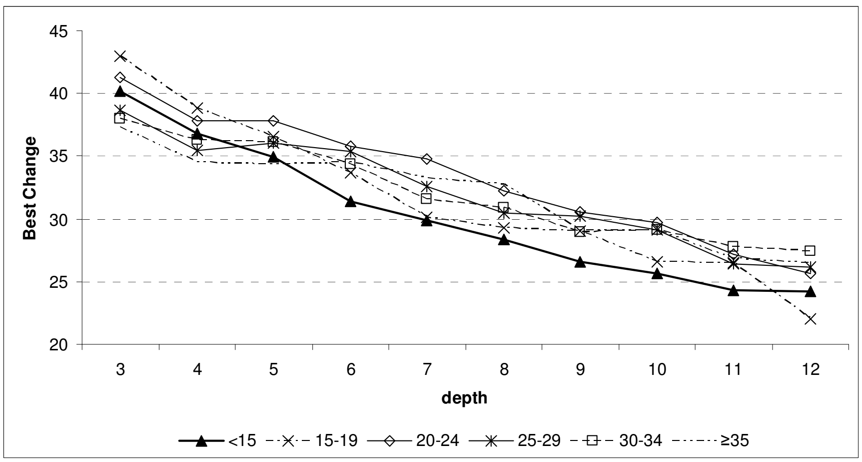

Figure 5: Go-deep results with positions with different amount of material on the board.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	37.33 (0.75)	100.00	-	-	0.07
4	34.53 (0.74)	74.93	25.07	-	0.07
5	34.39 (0.74)	64.85	27.02	8.13	0.07
6	34.53 (0.74)	56.06	27.96	10.77	0.07
7	33.29 (0.74)	50.55	31.48	8.11	0.07
8	32.78 (0.73)	44.96	31.90	9.42	0.07
9	29.13 (0.71)	43.32	30.05	10.02	0.07
10	29.13 (0.71)	40.65	31.72	9.85	0.07
11	26.80 (0.69)	35.84	32.76	11.34	0.07
12	26.53 (0.69)	31.71	35.20	11.27	0.07

Table 11. Results for the 4,112 positions of Group 6.

Search depth	Best Change in % (SE)	Fresh Best in %	(d-2) Best in %	(d-3) Best in %	mean evaluation
3	40.17 (0.86)	100.00	-	-	0.07
4	36.80 (0.85)	70.19	29.81	-	0.07
5	34.92 (0.84)	60.18	30.88	8.94	0.06
6	31.40 (0.82)	49.41	33.17	11.52	0.05
7	29.88 (0.80)	46.74	31.33	10.44	0.05
8	28.40 (0.79)	42.87	30.36	9.47	0.04
9	26.58 (0.78)	35.93	34.53	11.05	0.04
10	25.68 (0.77)	34.18	32.13	12.76	0.04
11	24.32 (0.75)	32.15	34.18	10.93	0.03
12	24.23 (0.75)	30.74	33.80	9.57	0.03

Table 12. Results for the 3,236 positions of Group 1.

The 95%-confidence bounds for Best Change at the highest level of search performed for the samples of 4,112 and 3,236 positions of Groups 6 and 1 are [25.20;27.90] and [22.78;25.73], respectively.

8 CONCLUSIONS

Deep-search behaviour and the phenomenon of diminishing returns for additional search effort have been studied by several researchers, whereby different results were obtained on the different datasets used in *go-deep* experiments. In this article we studied some factors that affect diminishing returns when searching more deeply. The results obtained on a large set of more than 40,000 positions from real chess games using programs CRAFTY, RYBKA, and SHREDDER suggest that diminishing returns depend on:

1. the values of the positions in the dataset,
2. the quality of the evaluation function of the program used,

and to some extent also on

3. the phase of the game, and the amount of material on the board.

Among other findings, the results also demonstrated with a high level of statistical confidence that both "Best Change" and "Fresh Best" rates (as defined by Newborn (1985) and Heinz (1998), respectively) decrease with increasing search depth in each of the subsets of the large dataset used in this study.

REFERENCES

Guid M. and Bratko I. (2006). Computer analysis of world chess champions. *ICGA Journal*, Vol. 29, No. 2, pp. 65-73.

Heinz, E. A. (1998). DARKTHOUGHT Goes Deep. *ICCA Journal*, Vol. 21, No. 4, pp. 228-244.

Hyatt, R. M. and Newborn, M. M. (1997). CRAFTY Goes Deep. *ICCA Journal*, Vol. 20, No. 2, pp. 79–86.

Junghanns, A., Schaeffer, J., Brockington, M. G., Björnsson, Y., and Marsland, T. A. (1997). Diminishing Returns for Additional Search in Chess. *Advances in Computer Chess 8* (eds. H. J. van den Herik and J.W. H. M. Uiterwijk), pp. 53–67, Universiteit Maastricht.

Newborn, M. M. (1985). A Hypothesis Concerning the Strength of Chess Programs. *ICCA Journal*, Vol. 8, No. 4, pp. 209-215.

Sadikov A. and Bratko I. (2006). Search versus knowledge revisited again. *Proceedings of the 5th International Conference on Computers and Games (CG2006).*

Schaeffer, J. (1986) Experiments in Search and Knowledge. Ph.D. thesis, University of Waterloo. Also printed as Technical Report (TR 86-12), Department of Computer Science, University of Alberta, Alberta, Canada.

Steenhuisen J. R. (2005). New results in deep-search behaviour. *ICGA Journal*, Vol. 28, No. 4, pp. 203-213.