

Abstract:

Aside from the historical origin like game theory, we want to show the important aspects that even now are of interest for us. It was started about 1922 by Bruno Rueger in Dresden.

And it was in common use before and after World-War II till to the mid 1970-ers. That was valid all over Europe at that time i.e. Germany, Austria, Yugoslavia, Netherlands, mainly.

The system is based on fundamental game-theory of the zero-sum game. It offers a ratio-scale starting with zero for the strongest play up to 80 for beginners.

Discussion about statistics shows the quality of rating by ratio-scaling against conventional ranking. The evaluation of tournaments prove the assumptions. There are considerations given about quantitative measuring of psychological behaviour.

I.) History

1a.) Theoretical Background

It is already some time ago (around 1880) that there was some activity and consideration in mathematics to work on the theory of games, and specially about the definition of a zero-sum-game. This is considered: chess and as well Baduk. Wiener and Shannon had given fundamental considerations on the cybernetics and about programming of such games. V. Neuman and Morgenstern (1944) gave fundamental aspects about applications to economics by game-theory. Boosted by such considerations and by the development of computing machines the programming of chess has begun. Progress however was getting on slower than was thought. But it is because we understand this now deeper and deeper, that the logics of game can be evolved by mathematical treatment to some amount with machines and vice-versa. (D. Levy 1988) This procedure however to-day is not completely solved and still implies severe problems.

Speaking of Baduk we may consider its exceptionally important role. For a long time it was believed and many people thought that it was too complicated to be treated by any logics of mathematical applied statistics and praxis. So it is only since ca. 25 years that some progress has been achieved here. So we may look on ESS, or better, we should see it in this connection of the beginning of applied mathematics and statistic. Yet, as well as it already does exist since the 20-er years of our last century, it fell rather almost forgotten, lately.

I b.) Beginning and the use of ESS.

Already when Baduk started to spread around Germany and Europe; (i.e. in D; Au; Yu; NI.) the countries with most of the players (a couple only,) it was 1922-24 that Bruno Rueger introduced a system of his own to rank the most often very weak playing people. He led the Go-Zeitung and by this he had the best survey about the few and dispersed living players receiving the paper. There was the big difficulty to give them a rating at all, as there were no meetings, nor tournaments at first! Maybe its number was around 100 in total! Only of a dozen of them an approximate playing strength was known. Many of the readers had never played together. - But the system was well accepted all over Europe

and thoroughly accepted by the experts. Here we remember in first place the past world-chess-champion Dr. Emmanuel Lasker who as well liked much to play Go-(Baduk). As Dr. of Mathematics he also intensified his study of games-theory. He explicitly was in use of ESS and accepted it, (1924 – 26). The system was commonly recognized.

Then in 1944 it was a little bit changed in order to ameliorate and adopt for more stronger and the weaker players. The ESS-system was in use till to ca: 1975 and it was normative for all Europe.

*) The names of : Go; Baduk; Weichi; hither-to will always be meant by „Baduk“ here.

However from the origin to these later days there was no publicly accepted or really known theoretical base of the system. Neither were there publications about it, - aside of course in the Go – Zeitung of Bruno Rueger. Why ? We don't know. It may be the destruction of many resources during and after the war, WW-2, or the exodus of many scientists by the Nazi-regime.

So, let us say suddenly and without any discussion it was abolished by the EgoF. in favour of the Japanese system. We may guess that the background of the ESS was not understood by the officials and that of course the Japanese system was known by the rest of the world and played by millions of people. Only in the Netherlands it has been taken up for some still longer time.

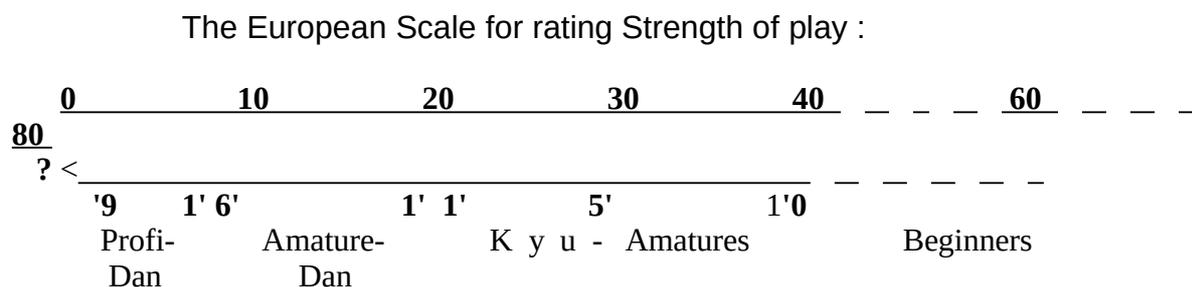
But it must not be forgotten, as apparently it has been-, that 1968 there was the publication of Dr. Walter Schmidt in the DgoZ. All the remarkable and essentially fundamental properties of ESS have been explained by him. His main object however was at that time like the international discussion, the amount of komi for the first move: to find out and to explain.- But aside of this, as he lived in Romania – Temesvar, and he had little contacts to evaluate many necessary statistical documents and data, he could not do further evaluations. So his results remained rather unprecise,- (but not uncertain!). Then in 1975 and 1979 the data have been intensely elaborated and added by diagrams by Dr.K.Heine. Another very exact mathematical- theoretical paper for elaborating the rating, rather than ranking, then has been given by Dr. W.Kramarczick 1979.

Summary:

There was as we may remark, a significant tradition for ESS. There does exist an important background by game-theory. However the system was left without discussion in favour of the Japanese traditional system.

II.) The System of ESS. - as it is:

a) ESS is a scale for rating the strength of Baduk-players from 0 ... to 80 (units). Here zero = 0 means the theoretical strongest way of play, as it is assumed that there will be no = zero mistakes for strongest optimal play. And 80 means the weakest play possible, let us say of a totally beginner, a totally ignorant player.



The traditional Japanese System for Ranking Strength of Play.

Fig. 1

The scale is orientated by the most possible amount of winning points i.e. ± 360 . But if both players perform the best way of play then the result will be ± 0 . For the normal human

player in between best and worst way of play the system of handicap stones is a useful help for compensation of big differences of playing-strength and high of victory.

The scaling itself is deliberate. As most convenient to be applied it was chosen to be $\frac{1}{2}$ handicapstone according to 5 (6) points of average victory high; to be the unit of the scale.

(But it is of no importance. Principally it could be taken as well to be: 1; $\frac{1}{4}$; or 2 – stones

of handicap.)

The scale itself, (not the scaling! -) i.e. to allot the strength of play to a rating, is fundamentally based, realizing, on game theory. Here it is the practical application to allot strength of play to the absolute criterium – height of victory (scores). It represents a mathematical corpus i.e. system that in itself has and possesses rigid order. It represents a complete logic structure.

b.) An example: For better understanding we may here explain an example from physics because it often may be difficult to understand that the above given statement for ESS - only represents a statistically given truth and in practice cannot be realized as „ein-eindeutig“ i.e. directly and clearly be proven. We often have this effect in progress of science. Just as we are used to look at things we figure to feel to feel about the procedure as good or as improper-bad.

The example here we take from the historical development of our knowledge of temperature: The pressure (p) as well the volume (v) of gases depends on the energy of its molecules. Thereby temperature is defined. Whereas in the beginning there were only expressions as „warm“ or „cold“ it was only around 200 years ago that a scale was introduced roughly working in the region of our human sensation.

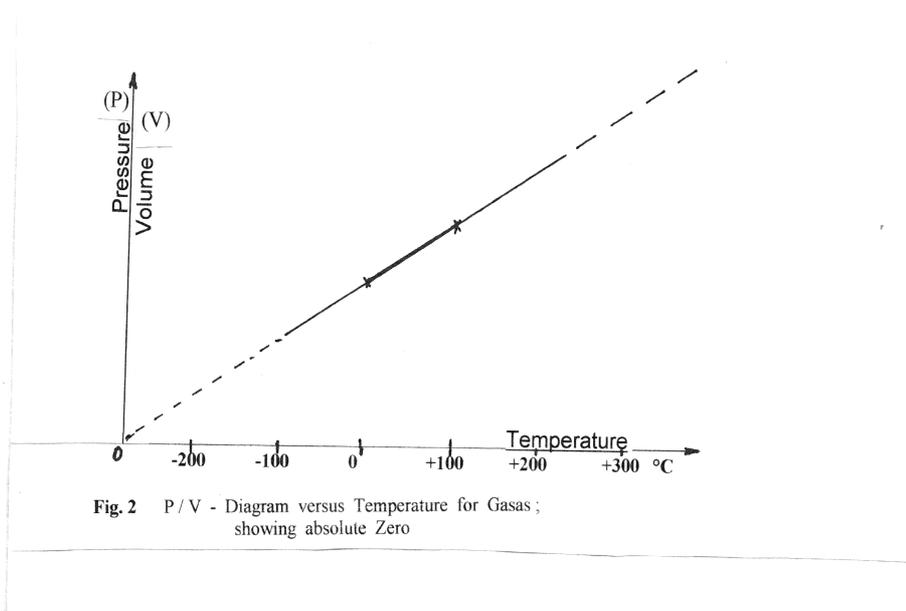


Fig. 2

This was the same for Celsius, Reaumur, Fahrenheit. Now it was possible to observe more precisely physical principles and rules. By separately developing the idea of energy-principles a theory about the absolute zero of temperature became evident. At first this only was a postulated hypothesis But before long, and more precise measurements, the postulate became to be a well established physical law. And now it is clearly meaningless as to what is the special units of the scaling: whether it is „C“; „Re“; „F“.- Essentially important only is the fact that there is a physical law established and thereby the system for itself can be considered as quantitative.

A similar development we do observe with ESS to more quantitative principles. Here it is: --> The scale, as based by score of victory essentially gives information about the quality of game. Thus it is a means of measuring logics and its rating. It is an objective measurement independent of psychology of human players.

c.) About reliability and precision.

Looking at the precision of statistical research we have to state, that there still remains a considerable width of error (uncertainty) of the results. So the situation of the beginning of the scale with zero still is uncertain by $\pm 0,5$ stone/handicap. A similar but bigger error may

be counted for scaling the strength for beginners with 80. But deficiencies like this should and shall not be taken for reasoning against the principium of installing such a scale.

Summary :

The ESS represents a continuous ??? scale of equally distant units. As it is based theoretically it rends ample possibilities for further research. The situation of the begin of the scale appears to be proven by statistics.

III) ESS. - Evaluation by hight of Victory :

The evaluation of tournaments of nearly equal strong players yealds new and interesting findings if always the complete scores of each victory or loss are taken into account.

The shown table may be an exemple:

European Championship 1975: Results as played: (: Black : White)

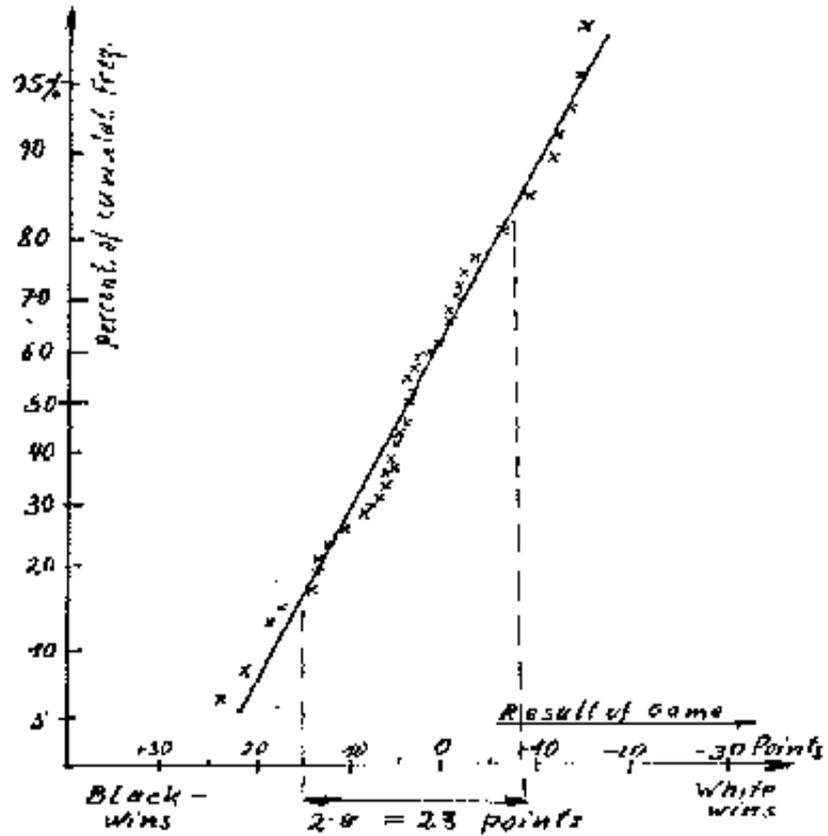
Nr	Name	KL.	1	2	3	4	5	6	7	8	9	10	11	12
1	Matern	9.0	≡	+a	+WO	—	+23,5	+a	+a	+12,5	—	+9,5	+a	—
2	Wimmer	10.2	-9	≡	+WO	+8,5	+a	+a	+4,5	+a	—	—	—	+a
3	Katscher	10.5	+8	-14	≡	-a	+13,5	+7,5	—	-0,5	—	—	+11,5	+5,5
4	Wiltsek	11.0	—	+4,5	+18	≡	-a	-a	+0,5	—	+33,5	—	+12,5	+a
5	Merisert	11.8	+9,5	-7	-6,5	+19	≡	+a	—	+2,5	+10,5	+a	—	—
6	Rehm	12.1	0	-6	+0,5	+20	-14	≡	—	—	+a	+8,5	—	+8,5
7	De Vries	12.1	0	+4,5	—	+4,5	—	—	≡	+8,5	+3,5	+9,5	-a	+7,5
8	Novak	12.6	-5,5	-3	+11,5	—	+1,5	—	-5,5	≡	+a	-10,5	+4,5	—
9	Greb	12.8	—	—	—	-20,5	-5,5	-11	+0,5	-14	≡	+a	+35,5	-5,5
10	Bates	13.4	+12	—	—	—	-7	-8,5	-1,5	+14,5	-12	≡	+4,5	-a
11	Kitsas	13.5	-7	—	+3,5	-3,5	—	—	+2,2	-2,5	-32,5	-4,5	≡	+a
12	Sudhoff	14.4	—	+6	-35,5	+2	—	+3,5	+4,5	—	+13,5	+20	-13	≡

Fig. 2. Results after corrections applied: (Resignation $\hat{=}$ 15 Point)

Fig. 3

We should consider that evolving scores and hight of victory the number of Komi for the first move for black must be respected. Games with resignation, if not too many, can as well be included, if a quantity of points will be taken as „punishment“ corresponding to the average scatter of the group-strength i.e. the collective.

Fig. 4



Now here can be recognized a remarkable fact: we see a function continuous connected between amount of scattering and the strength of players. (With our theoretical knowledge however this function no longer appears so much astonishing.)

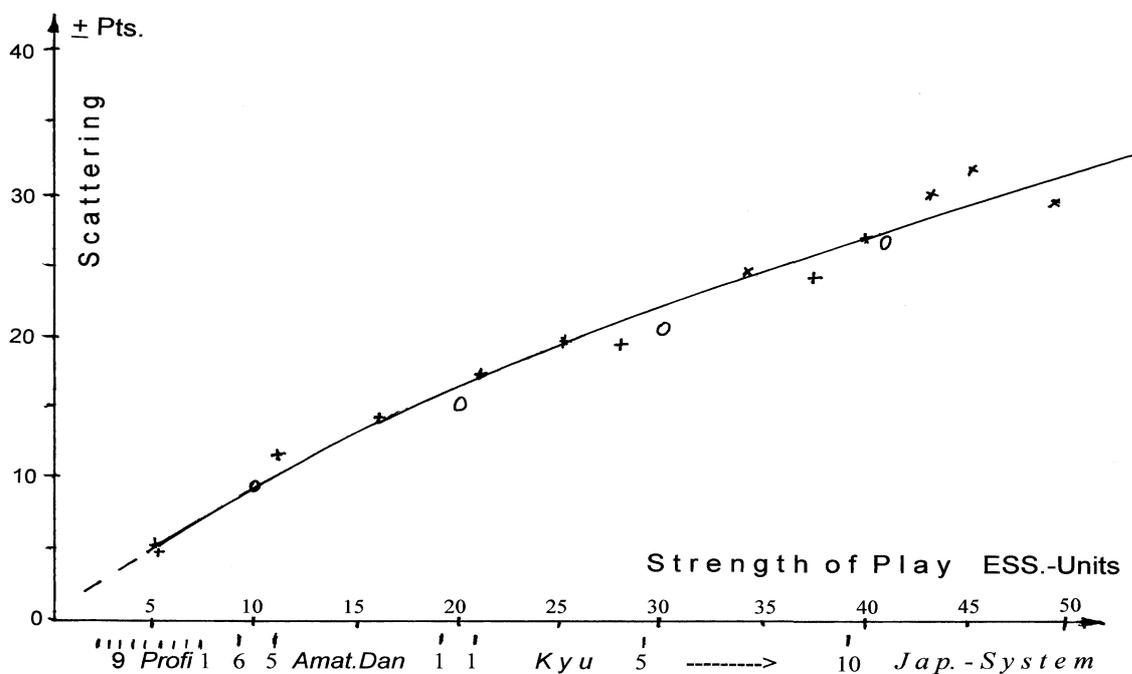


Fig.5. Function of Scattering versus Strength of Play

We see:

--> For better strength of play, the scatter will be less.

--> Thus it just means, the number and severeness of bad moves, mistakes, has been calibrated or measured by points.

--> We now clearly perceive the distance of the best now existing profi-players to the optimum theoretical play only is 1 handycapstone ($\pm 0,5$)

--> In ESS the hight of victory takes in important part, whereas up to now only the victory in itself is taken in account – indipendantly from the quality of play. Fundamental diffenences

in conception of play may result from such origin and as well in the way of evolving the results.

Another exemple here shall clarify the fundamental ideas:

If in 3-games once there is a win of 15 points, but two times a loss of only 1 or 2 points it should be obvious that the 15-pts.-win certainly is better than the $1+2 = 3$ pts. loss.

We see here, that the statistics with the scores of win and loss fundamentally is different than merely counting plus and minus for win and loss.

Similarly we further may register in 5-games: two wins by 8; 12 points to be better than three losses by 2; 3; 5 points.

Thus it should be accepted that criteria of significance in statistics must be applied . It is the inner logic of the game that each move should be optimum. If not, there are the mistakes deviating from optimum. The more and the worse they are, the more loss of points will count due to bad play. This should be respected when judging scores for the „value of game“.

As till now, it is normal praxis to prove strength of play by simply taking victory and defeat,

this is a system of ranking, where stones of handycap only represent a secondary scaling. However the manifestation of average scores of victory (ca: 12 pts. = 1 Stone Handcap) already implyes a law ruling kind of play-and-count.

In addition then it is the continuously observing of scattering of the results, and the continuous decline when better play is observed, it is the definite proof for the theoretically postulated relationship.

Summary:

It is possible to recognize the relation between strength of play and the amount of scattering of scores as a nearly linear function. So, when we elaborate game-results statistically, we distinctically do observe that the rating of an absolute optimal way of play is just $1 \pm 0,5$ stone of handicap away from the 9-Dan profi.-level.

We propose to evaluate the counting of victory in a way of statistic significance.

V.) Statistic – Classification: Ranking; Rating.

Sometimes it has been difficult to recognize the significant properties and dfferences of how to get the definitions of strength of play. Traditionally it is with most kinds of sports that some kind of ranking always is used. The reason for this is, because there are no other possibilities or procedures to compare :*better* or *not so good* of performance. With baduk however a sytem of rating is possible as there exists the inner law to yeald a fixed ratio scaling. But as it is difficult for understanding in praxis, it is still that evaluating of tournaments runs rather similar and the results do not be very diffenernt in general. Even if the system of rating promises to be more precise.

The differences of understanding are rather more by principle.

Ranking: means a mere definition of better or worse, respectively of victory or loss, respectively of something more or less. In many kinds of sports it is normal praxis to have or to use ranking: i.e: tennis, football, boxing, motorsports;.. and so on. This is a very useful method to get data or some procedure into some kind of order or to establish at least some relation amongst them. It must however be noted and not be forgotten, that there is by no means any inner function nor law necessarily implied.

Rating: Is the arrangement of data where there exists some rule, a law, by a scale with a definite origin = zero. Such a ratio-scale does access values. It is just a manifestation to look for a place where a new value (data, measurement) has to be placed on the scale.

For an example let us take the research on natural sciences where it is often such way, that first the phenomenons are to be classified and thereafter some ranking is applied to give some order for the arrangement. Then some hypothesis and a law are postulated. And they try whether all the data do comply to such a law and theory. Only, after all experiments are sufficiently satisfactorily positive we can establish a scaling, and the data can be put in for rating.

With both ways : ranking or rating (ESS) the evaluation of tournament results is done by theory and praxis of sample-taking. In both cases for the results there remains a certain amount of uncertainty due to the restricted size of sample. (i.e. the number of games for evaluation.) Only if there are an infinite= ∞ number of data we can expect exact results.

It must be stated however that a system of ranking always, and only, consists of a comparison.(Even if we do assume that there is some inner structure as in Baduk) On the other side the procedure of „ranking“ by a ratio-scale, like ESS, is a measuring, where the data are allotted to a given system. Thus the evaluated ratios yield considerably deeper scientific interpretation.

By first glance, apparently, this argument seems to be of quibbling-mode! (= a bad joke!) As in most cases the evaluating of games it does not make a big difference how we handle the results. Especially if there is the same procedure of sample-taking. But, however in principle, it does represent a fundamental difference.

Summary:

The definitions of ranking and rating have been explained. We tried to explain and to clarify that by ESS a broader and deeper evaluation is possible.

V.) Conclusion:

- ESS is based on principles of game-theory i.e.the definition of zero-sum game. As a ratio scale is defined from the beginning to the end, it appears to be nearly linear. Investigating in the game the possibility appears of measuring details. Thereby taking in account the scores, i.e. the height of victory is the most important tool. Scientific research and applications can be a big field of experience. As some example we may assume: behaviour of learning or evaluating the quantitative estimation and appreciation of positions in the game.

One outstanding result of ESS is the indication of the absolute zero of the scale, i.e. the allocation of optimal play to be: $1 \pm 0,5$ handicap-stones beyond the level of now best profi-players.

- There is a broadly and much acknowledged comprehension that the game (Baduk, Go, Weichi) represents a process of continuous optimisation. Finally this understanding includes that the height of victory is the most decisive factor in judging the quality of the game.

In this, such a manner we have to understand the many and intensely discussions about the rules. And it is of no use to have discussions about systems and kind of scaling, as long as the basic difference between rating and ranking is not understood. This has to be accepted by statistical considerations.

- Finally we may remark, the research up to now is quite encouraging even if we have to

admit: it is still rather unprecise by its statistical base. To get a better precision it will be necessary to have much more measurements. Let us say, effort and expenditure has to be enlarged vastly. It is no longer possible for one person to perform this by his own.

Bigger projects only can help.

- I want to express my personal conviction: Only by ESS there will be possible a deeper and broader public estimation aside from the insiders. It does certainly not concern the way of play, its outstanding high level, our sentiments. But it is the importance for cybernetics, understanding of complex systems and how important for us are statistic meanings, as our language does not contain statistic understanding.

Literature:

Bruno Ruger, *Ein Vergleich der Spielstarke japanischer und deutscher Gospieler*. Deutsche Go-Zeitung, 1922, Nr. 4, S. 1-3.

Bruno Ruger, *Einteilung der Leser nach ihrer Spielestarke in 60 Klassen*. Deutsche Go-Zeitung, 1944, S. 29-30.

John von Neuman, Oskar Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Walter Schmidt, *Statistische Abschatzung des Wertes des ersten Zuges beim Go-Spiel*. Deutsche Go-Zeitung, 1968, Nr. 1/2 S. 8.1-1-4, Nr. 3/4 S. 8.1-4-10.

Klaus Heine, *Statistical Research with the Game of Go*. In: "Proceedings of 2nd Sem. of Scientific Go-Theory" (K.Heine, Ed.), Mulheim a.d.Ruhr, 1980, S. 33-45.

Wolfgang Kramarczik, *Mistakes, Strength and probabilities of Winning*. In: "Proceedings of 2nd Seminar of Scientific Go-Theory" (K.Heine, Ed.), Mulheim a.d.Ruhr, 1980, S. 46-54.

David N.L.Levy, *Computer games-II. P87-ff*. Springer,N.Y. Press, 1988.

Franco Pratesi, *The Go Ranking System of Walther Shmidt*. British Go Journal, Nr 124, Autumn 2001, p. 42-43

Address: Prof.Dr.Klaus Heine, D- 26382-WHV; Peterstr.10a.

Tel: 04421-44526 Fax: 04421 7794766

Email: klaus.heine@ewetel.net

HP: <http://www.jerusoftware.de/klgeheine>