Naturwissenschaften 83, 272–275 (1996) © Springer-Verlag 1996

Dowsing Reviewed – the Effect Persists

H.-D. Betz, R. Kulzer
Sektion Physik der Universität, D-85748 München

H. L. König
Institut für Elektrophysik der Technischen Universität, D-80333 München

J. Tritschler
Institut für Medizin, Informatik und Systemforschung der GSF, D-85758 Neuherberg

H. Wagner
Institut für Pharmazeutische Biologie der Universität, D-80333 München

Further comment on J.T. Enright: Water Dowsing: the Scheunen Experiments. Naturwissenschaften *82*, 360 (1995)
On the next pages will appear a final reply by J.T. Enright.

We reply to a recent critique by Enright in *Naturwissenschaften* [1] on our *Scheunen experiment* performed as part of a research program on dowsing [2, 3]. Our contention had been that the size of the effect observed in this particular experiment was small but highly significant. Enright reviews our design and procedure in high terms (*the largest carefully controlled scientific study of dowsing ever conducted*), but he distrusts our mode of data analysis which, in his view, though correct, was *special, unconventional, and customized*, and replaces it by *more ordinary analyses*. The results of his treatment lead him to conclude that our *entire research outcome can reasonably be attributed to chance* and *if reason prevails* our study should be considered as the *last major study of this sort that will ever be undertaken.*
Ertel [4], however, subjecting the same data to sophisticated scrutiny, obtained results which entirely contradict those of Enright and even outperform our own original positive conclusion. The purpose of the present paper is to show why our former analysis of the data, backed by Ertel's confirmation, was correct, and why Enright's analysis failed. We reinforce Ertel's suggestion to expand and promote dowsing research in the future, on the basis not only of our experimental findings in the Scheunen experiment, but also of independent field investigations into onsite water dowsing of which some examples will be provided [5]. The entity of solid evidence presented by now no longer rationalizes calls for dismissal or the end of scientific discussions on the topic.
In the strictly controlled double-blind test series, a 10-cm diameter copper pipe (empty or with water flow) was randomly placed along a 10-m path in 1-dm steps. We ran 843 trials with 43 operators who had to locate the pipe along a test line on the floor above [2, 3]. Our original evaluation was performed with a preconceived multinomial distribution: success of a trial was defined by means of a Gaussian with width $\sigma = 0.5$ m, centered around the target position and digitized to seven ranges with nonzero weighing factors [1]. This procedure gives reasonable weight to both very close and less close hits, and appears sound with respect to previously observed operator behavior, where we noted that some scan the path "precisely", while others proceed more roughly and do not distinguish very small steps. All choices outside the defined ranges are misses with zero weight. Without dowsing effect the evaluation procedure would produce no result (null hypothesis). After completion of the thus-defined analysis, we checked whether our result depends critically on the prechosen σ; this was not the case, and we stated so in our final report. Thus, Enright's casting suspicion on our data analysis is unfounded.
All test series were analyzed independently and it proved to be that 13 of 107 series had chance probabilities below 5%. Since the probability of obtaining the actually observed number of significant (<5%) and nonsignificant series by pure chance was 0.07%, we termed the outcome highly significant (Table 1). At that time, no other kind of data treatment was attempted.
On what grounds could Enright come to entirely different conclusions? Apparently, his data analysis was too crude, and even illegitimate. His principal method of calculating correlation

Table 1. Number of observed and expected test series in the Scheunen experiment with chance probabilities α in three categories, obtained from our original multinomial analysis. The probability of obtaining the observed number of significant (≤5%) and insignificant series by pure chance is 0.7‰, and suggests a locating skill in the group

α [%]	n_{obs}	n_{exp}
$\alpha \leq 1$	3	1.07
$1 < \alpha \leq 5$	10	4.28
$5 < \alpha$	94	101.65

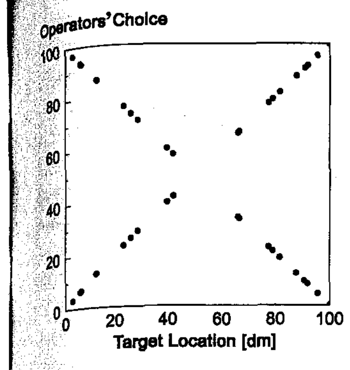

Fig. 1. Computer simulation of hypothetical data in Scheunen experiments: when operators' choices in 40 trials are either direct hits or locations symmetrical to the midpoint, the correlation coefficient is $R = 0$, and thus fails to detect an obvious correlation between target and chosen locations

coefficients, R, is particularly insensitive, and suffers from intrinsic shortcomings, because it is not based on the only parameter of importance, namely the distances between target and response, $Y(i)-X(i)$, but on terms $Y(i)-Y$ and $X(i)-X$ with averages X and Y. For example, let us assume the hypothetical case that operators either hit directly or chose a location symmetric to the midpoint (Fig. 1): one obtains $R = 0$, a completely insignificant result, though the response function is highly correlated to the target data. In fact, this mirror effect is not quite as hypothetical as may appear at first glance, since it proves to be inherent in the data [4]. Furthermore, R is undefined when constant choices $Y(i)$ are made, and R depends critically on the distribution of target locations $X(i)$ even when all distances $Y(i)-X(i)$ remain unchanged.
Inadequacy of the method is also demonstrated when we examine the data from dowser #99, to whom we will give particular prominence below: his 40 trials yield a correlation coefficient $R = 0.06$, which means total insignificance. In order to reach the 5% level of significance, a much better performance yielding $R > 0.31$ would have been required. However, 10 out of the 40 trials (25%) fall within 45 cm of the target, and a binomial evaluation testifies high significance (Fig. 2). It is evident that the correlation method weighs gross mislocations too heavily and honors close hits too weakly. For the academic Scheunen experiment, the quoted hit rate of some 25% must be considered high; a method of analysis which cannot detect an effect of this magnitude does not qualify for reliable conclusions.
Enright claims, without proof, that a simple binomial analysis gives no evidence for the dowsing hypothesis. In order to check this assertion, we carried out such calculations and found different evidence. Since the result is expected to depend on the chosen width of the hit interval, D, an analysis in retrospect must consider all possible and reasonable widths and cannot – as Enright's was – be limited to just two values. Figure 3 shows the result for the first 11 choices of D (± 0.5 to ± 10.5 dm). Interestingly, the observed number of hits lies constantly above chance expectation. The probability of attaining the respective result by chance depends indeed on D, and fluctuates around the 5% level of statistical significance, which is reached in four cases and barely missed in two cases. If there is a small dowsing effect in our data, the method is slightly too insensitive to detect it unambiguously. The reason is obvious: when D is chosen small, near misses are neglected, and when D is chosen large, close hits are not honored. In any case, Enright's conjecture that a binomial analysis shows total absence of any effect in our data does not appear as a wise or legitimate conclusion.
Since Enright singles out and comments on the performance of dowser #99, we show the corresponding results in Fig. 2. Here, the effect size is large enough for the binomial analysis to lead to an unequivocal result: for intervals with reasonable size the purely chance probability of attaining the observed number of hits is well below 1%. For example, 8 and 11 out of his 40 trials were closer than 25 and 55 cm to the target, respectively. This particular outcome must be termed highly significant.
Enright's most maladroit procedure concerns his handling of absolute average distances, $Z = |X(i) - Y(i)|$, between target and operator indications, which he applies to a selection of six "best" dowsers. He concludes from his calculation that *these dowsers could have performed better if they had – as*

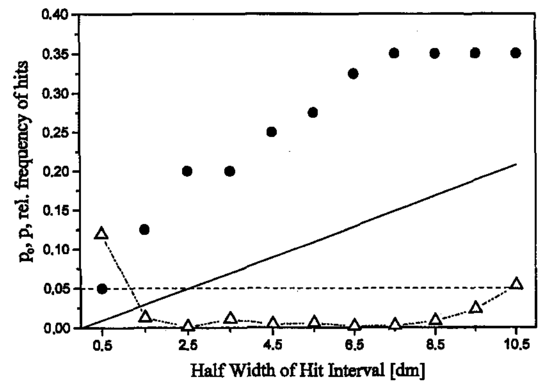

Fig. 2. Binomial analysis of the 40 trials by dowser #99 in the Scheunen experiment; chance probabilities p for the observed outcome (△; the *dotted line* serves to guide the eye) and relative frequency of hits (● observed; *solid line* chance expectation p_0), as a function of the width of the hit interval.
The *dashed line* identifies the 5% level of significance. The outcome is highly significant

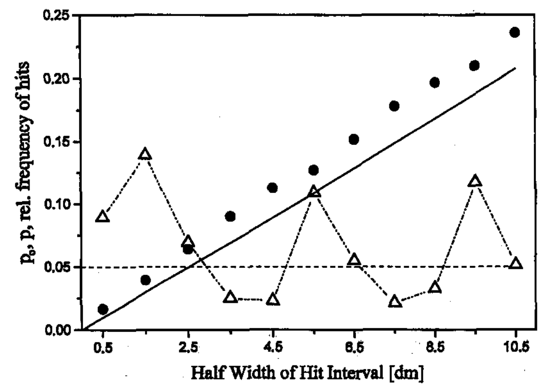

Fig. 3. Binomial analysis of the Scheunen experiment (all data without considering exclusion of small zones in some series [1, 2]); notations as in Fig. 2. Since the 5% level of significance is reached or approached in about half the cases, the outcome is ambiguous, but does not allow clear rejection of the dowsing hypothesis

successful strategy – simply chosen the Midpoint $Y(i) = L/2$ in each trial, instead of attempting to dowse. Here, Enright is in error. Even a nonexpert will readily realize that (1) a method does not qualify when it allows for a *successful strategy* in the absence of a real effect, and (2) dowsers to whom we – and to some extent even Enright – attribute above-chance performance cannot possibly do even better by choosing a trivial technique which, in itself, can produce only random results. The reason for the contradiction is obvious: Enright uses two different and incompatible *null hypotheses* at the same time. For constant midpoint choices, one obtains $Z_m = L/4$, but when dowsers make quite random choices over the entire path – and that is what they actually did – the pure chance expectation value to compare with is $Z_r = L/3$, rather than $L/4$, as supposed by Enright. Table 2 demonstrates that three of the six dowsers achieved better than chance results, excluding straightforward and certain denial of an effect. A thorough Z-analysis, which takes properly into account the actual distance *distribution* of all tests (not of a selected subset), has been worked out by Ertel and reveals not only highly significant dowsing skill, but also a most unexpected correlation between data and test site geometry [4]. According to our *null hypothesis,* all dowsers were considered as producing pure chance results, and different dowsers were allowed to contribute different numbers of trials. Due to this

Table 2. Distance analysis of the Scheunen experiment, according to Enright: average location error Z of six dowsers (the six best selected by Enright). N_{tot} number of trials. Since the operators placed their choices quite randomly over the 10-m path, the chance expectation value for Z is $L/3 = 3.33$ m; three of the six dowsers perform markedly better than by chance. This result does not allow clear rejection of the dowsing hypothesis

Operator	N_{tot}	Z [m]
# 18	78	3.42
# 99	40	2.93
# 108	60	3.48
# 23	129	3.64
# 89	23	2.45
# 110	20	2.80

Table 3. Evaluation of the four test series (with ten trials each) by dowser # 99 in the Scheunen experiment with a multinomial distribution; α is the probability of obtaining the observed results by pure chance. The significant result from series 2 is reproduced in series 3

Series	α [%]
1	8.5
2	0.2
3	3.2
4	18.0

concept, we could not rank among operators or examine individual repetition rates in the Scheunen experiment, as then further experiments with selected dowsers would have been necessary. Nevertheless, there is plenty of evidence for nonrandom repetition of success. First, we point to the performance of dowser #99: he was invited to participate as *favorite* for reasons described below, turned in the best series, and did repeat success (Table 3); dowser #110 was also specifically invited because he was known to have succeeded in a similar test before [6]; he proved to fall among those producing significant series in the Scheunen experiment. Second, Ertel's split-half and retest analysis shows that in the entire data hits and failures replicate [4]. Third, about 15% of the operators reproduced success in another test series not discussed here (*Laufbrett experiment* [2]). Finally, the most convincing evidence comes from unconventional water dowsing discussed below.
Ertel showed that a plot of operator responses along the test path with target position as parameter yields a set of curves which exhibit a regular target-dependent pattern. Aside from primary peaks of responses near target locations, the data show secondary peaks at mirrored target positions. In particular, and in pronounced contrast to random responses, the distance between real and mirrored target positions shrinks when the target is shifted successively from the end towards the middle of the path. It seems as if reflections due to the geometry of the roof above the test path were involved. The newly found effect is sufficiently prominent to warrant further attention and corresponding repetitions of the experiment: perhaps for the first time, it gives indications for a physical correlation to dowsing and renders theoretical onsets worthy of discussion [2]. In a quantitative analysis with randomization techniques, Ertel investigated the *distribution* of absolute distances between target and dowsers' choices: close hits prove to be clearly favored, and statistical significance reaches a high level of confidence ($p < 0.0001$).
Since Enright ignores essential and numerous published results of our more extensive dowsing research program [1, 7–10], a fundamental misunderstanding runs throughout his paper: although the Scheunen experiment was largely a search for an empty pipe, he identifies it as *water dowsing.* From the very beginning, and based on existing experience, we distinguished two intrinsically different situations for testing dowsers, namely locating (1) small *artificial* objects (e.g., pipes), and (2) *natural* subterranean discontinuities of arbitrary and not necessarily known nature (e.g., water-bearing fissures, dry faults, cavities). For this reason, the Scheunen experiment – whatever its result – cannot be used to assess actual water dowsing in a natural environment.
For these reasons, we continued with a new study, exclusively devoted to water dowsing. A detailed summary as of 1993 has been published [7, 10]. Two conclusions can be drawn which are based on independent data:

- artificial and small objects are generally difficult to locate; single trial hit rates are small and a dowsing effect is hard to prove;
- natural underground discontinuities of sufficient extension are readily located by a few, selected dowsers; they work sufficiently reproducibly to carry out this unconventional prospecting as a routine commercial profession.

Ample evidence exists for the outstanding performance of water dowsers [8]. It happened that dowser #99, who produced significant results in the Scheunen experiment, had been most successful with water dowsing in many foreign aid programs of the German Government. We have closely observed and thoroughly researched his continuing unconventional prospecting activities in arid areas of ten countries; specific scientific tasks were implemented, and the outcome was assessed by a committee composed of geoscience experts. For example, far more than 1000 drillings have shown that his average success rate is better than 80%, far ahead of that provided by modern conventional water-prospecting techniques in similar cases [7]. We like to stress that these examples, like other comparably valuable information, cannot be termed *anecdotal* and are to be taken seriously; extensive records exist and numerous experts have been involved in the studies, which have gone on for more than 8 years. Many similarities have been found in water-prospecting performance of modern technical devices and dowsers, which indicate a logical background and warrant continuing interest.
There are plenty of other highly successful dowsers who deserve thorough scrutiny. For example, we have investigated a drilling company which locates every well by dowsing techniques and gives a guarantee of success [7]. If the promised amount of water is not encountered, the client does not have to pay at all. The company has worked exceedingly successfully for over 10 years, and we can state from our surveys that no competing and conventionally operating companies can give such a guarantee, because it would soon drive them to ruin.
In the meanwhile, many critical scientists have been converted to accepting the facts [11], though they often claim that unquestionable successes result not from a particular dowsing skill but from prospecting experience. We discussed this partially valid argument for a long time before we published our detailed report [7]: too many cases defy a general explanation based on these grounds. Nevertheless, supposing that this is at least a partial solution of the water-dowsing puzzle, would it then not be worthwhile to better specify this kind of "experience", which allows water prospection in arid areas with two- to four times the success, at less than half the costs, and in a tenth of the time compared to conventional techniques? We consider this as a challenge for the future and suggest increased research efforts. Our newly presented results, the strange reflection effect in the Scheunen experiment and the striking success in water dowsing, seem to open new access to handling the still controversial questions, with the aim of attaining a long overdue scientific consensus which is not at variance with the observed facts, and, perhaps, may provide an understanding of the still persisting phenomenon.

1. Enright, J. T.: Naturwissenschaften *82,* 360 (1995)
2. König, H. L., Betz, H.-D.: Der Wünschelruten-Report – Wissenschaftlicher Untersuchungsbericht. München: Herold 1989
3. Wagner, H., Betz, H.-D., König, H. L.: Schlußbericht, BMFT 1990 (unpublished)
4. Ertel, S.: Naturwissenschaften *83,* 233 (1996)
5. A more detailed discussion of the topic will appear in a forthcoming issue of Journal of Scientific Exploration (1996)
6. Comunetti, A. M.: Experientia *34,* 889 (1978); *35,* 420 (1979)
7. Betz, H.-D.: Unconventional Water Detection – Field Test of the Dowsing Technique. GTZ Report, Eschborn 1993
8. Betz, H.-D.: Geheimnis Wünschelrute – Aberglaube und Wahrheit über Rutengänger und Erdstrahlen. Frankfurt: Umschau 1990
9. Betz, H.-D., König, H. L., in: Schulwissenschaft, Parawissenschaft, Pseudowissenschaft, S. 53 (G. Eberlein, Hrsg.). Stuttgart: Universitas 1991
10. Betz, H.-D.: J. Sci. Explor. *9,* 1, 159 (1995)
11. Raloff, J.: Sci. News *148,* 90 (1995)

Naturwissenschaften 83, 275–277 (1996) © Springer-Verlag 1996

Dowsers Lost in a Barn

J. T. Enright

Neurobiology Unit 0202,
Scripps Institution of Oceanography,
La Jolla, CA 92093, USA

Final reply to comments by S. Ertel [1] and H.-D. Betz et al. [2] on J. T. Enright: Water Dowsing: the Scheunen Experiments. Naturwissenschaften *82,* 360 (1995)

Very simple graphs of the Scheunen experimental results (plots of dowsers' choices vs. water-pipe locations), when thoughtfully considered, clearly demonstrate that dowsers with their various kinds of witching sticks usually lost their way completely, when seeking a hidden pipe in a barn (Figs. 1 to 4 in [3]). Attempts to discredit that graphically obvious conclusion [1, 2] demonstrate that statisticians who search through data, armed with various fancy tests rather than divining rods, can also lose their bearings.

The Original Data, without Statistical Cosmetics

Anyone who is interested in dowsing and the outcome of the Scheunen experiments should consult the graphs of the results from the six "best" dowsers that are presented as Figs. 1–4 in my review of the experimental data [3]. Those simple plots of the observed dowsers' choices relative to location of the hidden pipe stand on their own, independent of all statistical analyses. For example, it is an empirical *fact,* and not the outcome of testing some obscure null hypothesis, that five of the six "best" dowsers (including the famed #99) could have performed better on average if they had simply chosen the midpoint of the test line in each and every trial. It is difficult to imagine results that look more scattered than the