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 very 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

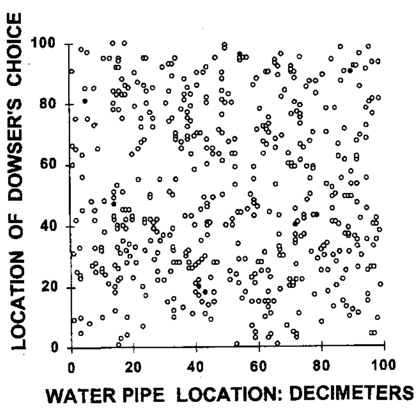

Fig. 1. Supplement to Fig. 1 in [3]: dowser choices plotted against water pipe location for 522 tests from the 37 dowsers who were not, according to [4], among the 6 "best". (Filled circles represent two points at identical coordinates). These data were also used by Ertel in [1], and were kindly supplied by him. As in Fig. 1 of [3], one can vaguely discern gaps and snakelike patterns in the graph, somewhat like recognizing constellations in the starry sky. They are probably due, in part, to the strikingly nonrandom distribution of pipe settings, which were selected using the apparently defective random generator of a computer [4]. Successful dowsing would be indicated by a concentration in the data comparable with the Milky Way, stretching from lower left to upper right. It is difficult to decide whether the data here are more scattered than the "best-dowser" data in Fig. 1 of [3]

summaries that are plotted in [3] – and those data were chosen for illustration because they came from the "best" dowsers, based on criteria of the investigators themselves [4]. In my opinion, such graphs argue so strongly against a significant dowsing effect ("significant" in its broader meaning, rather than its statistical usage) that a contrary interpretation could only suggest a compulsion to cling tenaciously to a failed hypothesis: nothing less than the will to believe.

The long and the short of it is that dowsing performance in the Scheunen experiments was not reproducible. It was not reproducible interindividually: from a pool of some 500 self-proclaimed dowsers, the researchers selected for their critical experiments 43 candidates whom they considered most promising on the basis of preliminary testing; but the investigators themselves ended up being impressed with only a few of the performances of only a small handful from that select group. Even more troublesome for the hypothesis, dowsing performance was not reproducible intraindividually: those few dowsers, who on one occasion or another seemed to do relatively well, were in their other comparable test series usually no more successful than the rest of the "unskilled" dowsers (see Figs. 2 and 3 in [3]).

Statistical Analysis

Perhaps the most important caveat about statistics that was emphasized in my review [3] is that proper application of the logic of inferential statistics demands that the test to be applied to a set of results be selected *before* any of the results have been examined. I expressed this concern about the original analysis [4] as follows: "The statistical procedure used in the Final Report for the Scheunen experiments is a special, unconventional, and customized analysis, and the report does not indicate whether this choice of statistical procedures was made before any of the critical experiments was performed" [3]. This issue is important because if the chosen statistical procedure was developed or selected with some of the data already in hand, the "probabilities" calculated and the "levels of significance" would lose all objective meaning. As I carefully emphasized, identical reservations apply to the more conventional statistical examination of the data that I myself undertook: with after-the-fact analysis, as were mine (and as are all the more recent analyses of Ertel [1]), no confidence whatever can be placed in probabilities derived from ordinary statistical testing; and I explicitly refrained [3] from drawing any inferential conclusions based on "statistical significance" from my own analyses.

The commentary by Betz, König, Kulzer, Tritschler, and Wagner [2] provided an ideal opportunity for them to totally disarm such concerns, by asserting that the complex multinomial test used in the original analysis was fully developed and had been chosen for application to those data before any of the critical experiments was conducted, i.e., prior to 9 April 1987. The absence of such a statement in their commentary may represent no more than a simple (but important and alarming) oversight; but if such an assertion cannot honestly be made, then statistical analyses of the Scheunen experiments in the Final Report [4], all of which were based on that test, are no more than an "exploratory" application of statistics, in which probability values are the outcome of an empty formalism.

That distinction is by no means a subtle issue. Exploratory analysis can be compared with using today's newspaper to find a scheme that could, in principle, have "predicted" the outcome of yesterday's horse races or yesterday's lottery: not genuine prediction but "postdiction", with no risk whatever that the wager can be lost. The probabilities associated with proper statistical testing are, instead, applicable only to genuine prediction, made before the race has been run or the lottery numbers drawn. The meaninglessness of "probabilities" calculated from inappropriate statistical "testing" applies with great emphasis to the reexamination of the Scheunen data described by Ertel [1], in which exploratory statistical analysis was pushed to an astonishing extreme. Some of the "hypotheses" that were derived from the data are so bizarre that no sensible dowsing enthusiast could have dreamed of suggesting them until after carefully sifting through the experimental data, and manipulating them in dozens of ways. This was followed by multiple testing, driven to the *n*th degree, so as to find the most "significant" way of reorganizing the observations. Yet that article [1] is fairly peppered with probability values and the terminology of inferential testing, leaving the impression that exploratory data analysis (i.e., "postdiction") deserves the same sort of credence and respect for purported "significance levels" that are warranted by the real predictions of legitimate statistical testing! Given enough freedom in the design of customized tests, the imaginative use of exploratory statistics *guarantees* that any data set can be made to seem nonrandom in several different ways. In the present context, such an outcome has nothing to do with the skill of the dowsers, as Betz et al. [2] and Ertel [1] contend; instead, it simply demonstrates the persistence and ingenuity of the statisticians. Perhaps the most remarkable aspect of this situation is that statisticians who are so adept at manipulating data seem to have lost their way at the outset, by failing to recognize the fork in the road between exploratory data analysis and genuine statistical testing.

As can easily happen in such after-the-fact analyses, Ertel [1] and the original researchers [4] used the same data to reach qualitatively different sorts of interpretations, based on different criteria – although they seem not to have noticed that discrepancy [1, 2]. Betz and colleagues [4] concluded that nearly all of their candidates showed no dowsing skill whatever; they decided, instead, that only a very few of the dowsers had (occasionally) shown noteworthy skill – where skill is (quite reasonably) demonstrated by proximity between hidden pipe and dowser's choice. Ertel [1] ignored those interpretations and, instead, charged into the data with the assumption that if dowsing skill exists, it would be equitably apportioned among all dowsers at all times – where skill can also be demonstrated by a "wrong" choice that is located at a mirror image of where the sought-after pipe was placed, reflected around a point that was 34 dm (yes, 34 dm!) from one end of the 10-m test line. The open-minded, thoughtful reader who takes a careful look at the simple plots of the data presented in my review [3; see also Fig. 1 here] will, I think, reach a still different conclusion: the recognition that the experiments led only to extremely scattered and unreproducible results, with no persuasive evidence at all for the existence of practical dowsing skill – no matter how persistent and ingenious the statistician may be who chooses to dissect, manipulate, and massage the data.

Can Reason Prevail?

I stand by my conviction: "...the Scheunen experiments are not only the most extensive and careful scientific study of the dowsing problem ever attempted, but – if reason prevails – they probably also represent the last major study of this sort that will ever be undertaken" (p. 369 in [3]). To my regret, however, I am not, at this time, completely convinced that reason will prevail.

1. Ertel, S.: Naturwissenschaften *83,* 233 (1996)
2. Betz, H.-D., König, H.L., Kulzer, R., Tritschler, J., Wagner, H.: ibid. *83,* 272 (1996)
3. Enright, J.T.: ibid. *82,* 360 (1995)
4. Wagner, H., Betz, H.-D., König, H.L.: Schlußbericht 01 KB 8602, BMFT 1990

Naturwissenschaften 83, 277–279 (1996) © Springer-Verlag 1996

Acetylation of *α*-Tocopherol by the Squash Beetle, *Epilachna borealis*

Defense Mechanisms of Arthropods, 134 [1]

A.B. Attygalle, J. Meinwald

Department of Chemistry, Cornell University, Ithaca, New York 14853, USA

C. Rossini, T. Eisner

Section of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA

The phenolic, alkylated benzodihydropyrans known as tocopherols are widely distributed in the plant kingdom [2], and have occasionally been reported from animal sources as well [3]. In addition, very large quantities of synthetic tocopherol esters (acetates) are produced industrially and used as antioxidants and dietary supplements in the food, cosmetic, and pharmaceutical industries [4]. In fact, the quantities of tocopheryl acetates found in municipal waste have been taken as an indicator of the domestic use of tocopheryl acetate-containing products [5]. We have found that tocopheryl acetates also occur in nature, specifically in the pupal exocrine secretion of the squash beetle, *Epilachna borealis* [6]. It seemed likely that these beetles biosynthesize these acetates from tocopherols found in their larval diet, and we now report the results of an isotopic tracer experiment that confirms this hypothesis.

One hundred microliters of a solution of 5,7-*bis*(trideuteromethyl)-*α*-tocopherol (*1*) in ether (1 mg/ml) was applied to individual leaves of zucchini squash plants (*Cucurbita pepo,* variety Milano). Following evaporation of the solvent, one each of such leaves was offered to six individually caged fourth-instar larvae (14- to 21-day-old) of *E. borealis.* This procedure was repeated with each larva for another 2–3 consecutive days until the larvae transformed into prepupae. Control larvae received either untreated leaves ($N = 6$ larvae) or solvent-treated leaves ($N = 6$ larvae). Larvae of all three categories consumed parts of the leaves. Six days after pupation, the secretion from glandular hairs of individual pupae was collected with microcapillaries (by the technique previously used [6]) and extracted into hexane (10 µl per sample). The samples from each pupa were analyzed by gas chromatography using a Hewlett Packard 5890 instrument equipped with a split/splitless injector and a flame ionization detector. The HP-ChemStation software program was used to acquire and integrate data. Solvent extracts were introduced by splitless injection. Low-resolution electron-ionization mass spectra were obtained using an HP 5890 gas chromatograph linked to an HP mass selective detector (MSD).