Consequently, after having adopted the new habit of building cells with plant hairs, selection might have forced the *Anthidium* bees to protect cell contents in an alternative manner. Extrafloral trichome exudates transferred to the nest wool may compensate for the protection exerted by resin. This view is supported by preliminary experiments, which demonstrated that plant wool soaked with glandular secretions exhibits increased water repellency, as well as by the ecological function attributed to the trichome exudates. These apparently serve either as contact poisons or as volatile repellents, inhibiting microbial infection of the plant surface and preventing attack by herbivorous arthropods [11]. We therefore hypothesize that the plant glandular secretions on the nest wool serve to waterproof the brood cell, to prevent microbial attack of provisions and larvae, and/or to deter nest-robbing arthropods. In addition, they may facilitate the manipulation of the plant hairs by glueing the wool threads together. However, a few *Anthidium* species and the representatives of the genus *Pseudoanthidium* which also build their brood cells with plant hairs [8, 12] are not equipped with specialized tarsal brushes for harvesting plant glandular products. This shows that the collection of plant hairs for cell construction is not necessarily connected with nest wool impregnation in every case.

We thank M. Endress for linguistic help, C. Westerkamp and an anonymous referee for helpful comments, and U. Jauch and A. Zuppiger for technical assistance. This work was supported by a grant of the Ciba-Geigy-Jubiläums-Stiftung.

Received September 25, 1995 and January 10, 1996

1. Michener, C.D., Griswold, T.L.: Univ. Kans. Sci. Bull. *55*, 299 (1994)
2. Vogel, S.: Trop. Subtrop. Pfl.welt *7*, 1 (1974)
3. Vogel, S.: ibid. *54*, 1 (1986)
4. Neff, J.L., Simpson, B.B.: J. Kans. Entomol. Soc. *54*, 95 (1981)
5. Whitten, W.M., Allen, M.Y., Williams, N.H.: J. Chem. Ecol. *15*, 1285 (1989)
6. Cane, J.H., Carlson, R.G.: Comp. Biochem. Physiol. *78B*, 769 (1984)
7. Altenkirch, G.: Zool. Beitr. *7*, 161 (1962)
8. Pasteels, J.J.: Ann. Soc. Ent. Fr. *13*, 651 (1977)
9. Müller, A.: Ecol. Monogr. (in press)
10. Ghisalberti, E.L.: Bee World *60*, 59 (1979)
11. Hafitz, A.: Phytopathology *42*, 422 (1952); Levin, D.A.: Quart. Rev. Biol. *48*, 3 (1973); Levin, D.A.: Ann. Rev. Ecol. Syst. *7*, 121 (1976); Craig, R., Mumma, R.O., Gerhold, D.L., Winner, B.L., Snetsinger, R., in: Natural Resistance of Plants to Pests: Role of Allelochemicals, p. 168 (M.B. Green, P.A. Hedin, eds.). Washington: Am. Chem. Soc. Symp. Ser. 296, 1986
12. Westrich, P.: Die Wildbienen Baden-Württembergs. Stuttgart: Ulmer 1989
13. Swofford, D.L.: Phenetic Analysis Using Parsimony (PAUP). Version 3.1.1. Champaign: Illinois Nat. Hist. Surv. 1993
14. Maddison, W.P., Maddison, D.R.: MacClade: Analysis of Phylogeny and Character Evolution. Version 3.0. Sunderland: Sinauer 1992

Naturwissenschaften 83, 232–235 (1996) © Springer-Verlag 1996

The Dowsing Data Defy Enright's Unfavorable Verdict

S. Ertel

Institut für Psychologie der Universität, D-37073 Göttingen, Germany

Comment on J.T. Enright: Water Dowsing: the Scheunen Experiments. Naturwissenschaften *82*, 360 (1995)
In the next two issues there will be a further comment by H.-D. Betz et al. and a final reply by J.T. Enright.

König and Betz investigated the skill of dowsers to identify the position of a hidden pipe, the experiment was performed in a barn (Scheune). Based upon an analysis of their Scheunen data, the authors contended that the empirical evidence for such skills is compelling [1]; Enright [2], however, summarizing his reanalysis of the same data concluded:

"No persuasive evidence was obtained for a genuine reproducible dowsing skill. The absence of reproducibility suggests that the entire research outcome can reasonably be attributed to chance."

In order to help resolve this contradiction, the present author also reanalyzed the complete data set. Dowsing success was assessed by the *absolute distance* on the test line between the dowser's response[1] and the location above the hidden pipe. Enright analyzed the Scheunen data by spatial correlations, not considering that the direction towards which a dowser's response deviates from the target (plus and minus) should be ignored. König and Betz, utilizing multinomial frequency counts, correctly considered only the size of response deviations ignoring irrelevant directions.

Next, the definition of an exact hit is arbitrary and unnecessary for an analysis: Enright assumed that effects, if existent, must be precise enough to become visible by glancing at spatial scatter plots. König and Betz were more careful by allowing for a hit tolerance of ca. 10 dm. Nevertheless, they, too, excluded valuable information, as will be shown later. We sought relationships between stimulus and response without preconceived hit definitions in order to avoid a Type B error, i.e., erroneously concluding that an effect does not exist.

Third, our analysis is based on all trials of all dowsers ($N = 843$), not on select-ed samples of dowsers (Enright) and not on runs of trials within individual sessions (König and Betz). Advantage: statistical degrees of freedom (df) are greater for all trials taken together increasing the precision of analysis. Legitimacy: target-response relationships for individual trials are stochastically independent of one another, even within one dowser's session, since target locations were randomly determined.

Fourth, expected distances between responses and targets ("expected" = "expected by chance"), indispensible for assessing error probabilities of observed responses, are obtained by *randomizations* (to be explained below)[2]. Enright did not consider expectancies at all and König and Betz' assumptions underlying their use of multinominal distributions are not as certain as the present procedure. Expectancies by randomization help avoiding a Type A error: erroneously concluding that an effect exists [3].

A first check. *Expected* distances were obtained as follows: 843 observed dowser responses, varying between 1, 2, and 100 dm, were paired with the original 843 target positions randomized, i.e., the positions were shuffled, so that, e.g., the dowser's response no. 1 in the series of 843 trials which had occurred at, say, 10 dm on the line came to be paired with pipe position no. 515 associated with, say, position 60 dm. The absolute distance between response and target is 50 dm in this case. After calculating absolute distances between 843 dowsers' responses and randomly paired pipe positions, the latter were shuffled again to be randomly paired with the dowsers' responses, another 843 distances thus being obtained. This process was repeated 1000 times yielding a total of 1000×843 distances, ranging between distance 0 dm (exact hit) and 100 dm (complete miss). The average occurrence of distances 0, 1, 2, 3,... 100 dm, i.e., each occurrence divided by 1000, defined 101 expectancy levels (E = expected).

The *observed* values (O) were obtained by counting the distances between 843 dowsers' responses and 843 *unrandomized* pipe targets at original experimental locations. These counts were then related to the average counts for each distance with expected target locations by subtraction ($O-E$). The obtained differences $O-E$ are shown in Fig. 1. It can be seen that the greatest positive deviation from expectancy occurs at 0 dm above the pipe's hidden location. Moreover, the rise of FFT-smoothed deviations from right to the left culminating at 0 dm suggests that the pipe did, in fact, function as attractor (here called attractor I).[3]

There are three additional peaks along the test line, as if other attractors had been active in the barn; but what else, aside from the experimental pipe stimulus, could have attracted the dowsers' rods? A closer look at this is required.

An unexpected pattern of responses emerges by analyzing the dowsers' response frequencies separately for ten groups of target-response distances 0–10 dm, 11–20 dm,... 91–100 dm. Observed frequencies for the first five target positions are given in Fig. 2 (original frequencies, not their deviations from expectancy). Strangely, in eight of the ten frequency distributions, two peaks appear instead of only one, and the corresponding average distributions based on 1000 target randomizations do not display secondary peaks (not shown in Fig. 2). Another notable observation is that the distance between the two peaks diminishes as the target's location approaches the midpoint of the dowser's test line (31–50 dm).

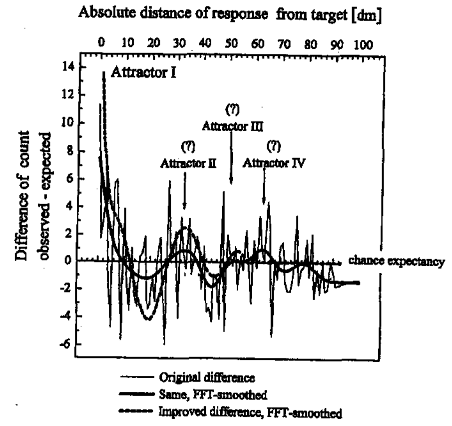

Fig. 1. Main results. Difference of observed response counts from chance expectancy, for all experimental trials ($N = 843$), occurring at 0, 1, 2,...100 dm absolute distance for the pipe's location. *Solid* 1st stage of analysis; *bold* FFT-smoothed results; *dashed* 2nd and improved stage of analysis (to be described below). Counts are based on a mirror-transformed scale (mirror axis at 34 dm). Note for the *dashed* curve: The second peak represents attractor IV, not attractor II, since attractor II effects have been joined here with attractor I effects

This oddity reminded me of one of Professor Betz's personal communications. It had appeared to him that the dowsers sometimes responded at locations opposite to where the pipe had been placed (a "symmetry error", as it were). He tentatively explained such responses – if they were real – as possibly due to the slanted roof of the barn whose ridge, projected down to the test line, would cross it about halfway. *Reflections of rays* (if rays were the transmitters) had to be stipulated in that case. In order to avoid generalized doubts in his dowsing study, Betz had omitted this casual and uncertain observation in publications. Interestingly, even Enright's visual inspection of his spatial scatter plot indicated this symmetry effect: "...quite a few points are also distributed along the opposite, downwardly directed diagonal".[4]

If such reflections occurred, then the data might also tell how they worked. For this purpose the test line was "mir-

[1] "Response", as used here, does not presuppose an empirical relationship to the experimental stimulus.

[2] The mathematical rationale of randomization procedures "is rather different to the classical or normal theory models...The probability statements made under randomization models do not refer to distributions over all samples, but rather to distributions over all possible randomizations of the same sample" [3]. Randomization, when applicable, is preferable to classical procedures ("the fairest test of all", [3]), and less demanding, the only requirement being appropriate computer hard- and software (programs have been customized here by SC, Statistical Calculator, MOLE Software).

[3] Frequencies of response deviations from target at scale position 0 dm (exact hits) were doubled for both expected and observed samples, in order to adjust their probabilities to deviations >0 dm, e.g. absolute deviations of responses of size 1 combine response deviations at −1 dm and +1 dm while exact hit occurrences (deviation = 0 dm) stand alone. The doubling of 0 dm deviations adjusts them to (−1 dm)+(+1 dm) deviations, i.e., to the probability level of all >0 dm deviations.

[4] Enright, p. 364, see also his Fig. 2A. After removing all points at the main upward regression line, a regression for the remaining points will have the opposite, downward direction.

ror-transformed".[5] Dowsing responses that occurred opposite to the target's location ("opposite" defined by a mirror axis) were thus treated as if they had occurred on this side of the mirror axis. Mirrored locations of pipe or response are thereby treated like real locations.

If mirrored responses occurred, then the question arises where the axis of symmetry was located. The computer therefore searched, along the original test line between section 20 dm and 80 dm, the axis of symmetry yielding maximum effect size. "Effect size", defined by Cohen's kappa [5], evaluates numerical differences between observed and expected frequencies.[6] This search would yield between dm 20 and 80 a random kappa distribution if no mirror effect existed; a conspicuously peaked curve must be expected if mirrored responses did occur.

Searching an effect size maximum went hand in hand with searching for another empirical maximum: at which hit tolerance will kappa culminate? A hit may be defined as a response at distance 0 dm from the target, or at distances ≤ 1, ≤ 2, ≤ 3 dm, etc. kappas were calculated for 0 dm through ≤ 25 dm (= 26 integer positions), separately for each mirror axis.

Results (see Fig. 3): first, kappa peaks with hit tolerance ≤ 11 dm; second, among the axes of symmetry the axis at 34 dm reveals the greatest effect. Consequence: subsequent data analyses ought to be based on hit tolerance ≤ 11 dm and on an axis of symmetry at 34 dm ("mirror axis").

For secondary peaks 54–56 dm and 68–70 dm, kappas are also conspicuous. That is, even attractors III and IV, suggested but still questionable in Fig. 2, become more pronounced.

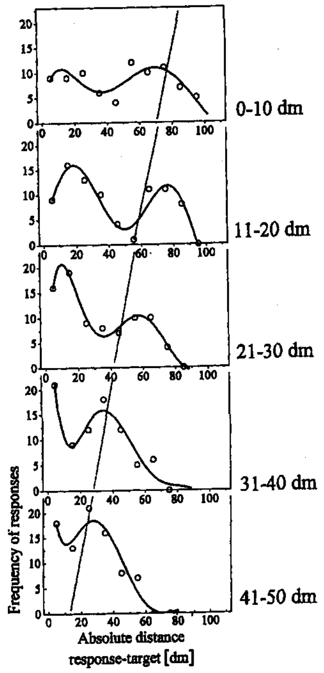

Fig. 2. Strange secondary peaks. Counts of dowsing responses broken down for the first five of ten target locations. Original counts are shown here, not deviations from expectancy, as was the case in Fig. 1. Note: absolute response distances decline for targets > 50 dm with target locations approaching the midpoint (41–50 dm) of the test line. This is a procedural constraint, no result, since the greatest possible deviation from a target of a dowsing response declines when the target approaches the midpoint. The main result is that secondary peaks interrupt the declining frequency curves. The *slant of the dotted line* across secondary peaks shows that their distance from the main peaks (*outer left*) diminished as the targets approached the midpoint of the test line (range 41–50 dm)

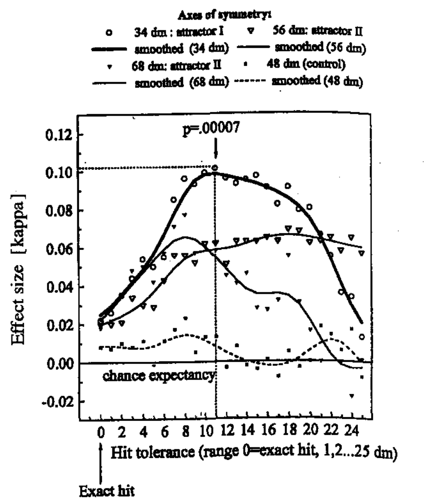

Fig. 3. Maximizing kappa (2): kappas (*vertical*) for varying hit tolerance definitions (*horizontal*) and mirror axes 20 dm through 80 dm. Of 61 resulting curves, the three most conspicuous (effect-indicating) have been selected (another curve lacking any effect is given as a control case). Note: the control curve (*dashed*) shows kappas obtained by using a mirror axis of 48 dm which is apparently incapable of reflecting dowsing signals: the dowsing effect peaks with mirror axis at 34 dm using response tolerance ≤ 11 dm from targets

A finel significance test. By randomizing target positions with the mirror axis set at 34 dm and hit tolerance at ≤ 11 dm, following the above results, the observed effect reaches the largest significance ($p = 0.00007$). In other words, by shuffling the 843 target positions prior to calculating their distances from 843 responses, the observed number of hits occurs only 7 times among 100000 such runs.

Our understanding of what actually happened in the barn may be summarized as follows: dowsing stimuli existed, but they were deflected to some degree. König and Betz took into account an attractor I effect only and the researchers' significance level ($p = 0.007$) was conservative. Fortunately, the pipe stimulus did not entirely vanish by dissipation, some portion seems to have been reflected by slanted constructions above the test line down to nonpipe locations. Our analysis uncovered those secondary effects. As attractor II effects were joined with attractor I effects, the effect total was raised to a significance level ($p = 0.00007$), which can hardly be dismissed as not convincing (see dashed line in Fig. 2).

[5] An original linear scale A (the dowser's testing line) is transformed into a bidirectional symmetrical scale B using an arbitrary axis of symmetry (= mirror axis). An axis at 30 dm, for example, transforms A into B as follows:

A	10	20	30	40	50	60	70	80	90	100	dm
B	20	10	0	10	20	30	40	50	60	70	dm

After scale transformation the pipe's location at 10 dm in A is equivalent to its location at 50 dm in A. In general, any target or response on one side of the bidirectional line B is equivalent to the respective target or response on the opposite side of that line, "opposite" with respect to the selected mirror axis. The presupposition underlying this procedure (true with approximation only) is that the energy of rays (or whatever it is) is not diminished through reflection by the barn's slanted roof. The mirror axis effective in the Scheunen experiment will be obtained by kappa maximizing calculations (see main text).

[6] Kappa = (observed frequency for a critical condition minus expected frequency for the critical condition) divided by (the total of observed frequencies minus expected frequencies for the critical condition).
Example A: 15 critical observations, expectancy = 10, $N = 100$ total observations.
Example B: 15 critical observations, expectancy = 10, $N = 1000$ total observations.
Kappa (A) = (15 – 10)/(100 – 10) = 0.056.
Kappa (B) = (15 – 10)/(1000 – 100) = 0.0056.

Table 1. Frequency of hits (with hit tolerance ≤ 11 dm) for test and retest conditions, ten trials each, based on a mirror-transformed scale of the test line at 34 dm. (Examples: subject no. 10 participated two times, no. 18 six times, no. 108 five times)

Pair no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Subject no.	10	18	18	18	23	23	23	23	23	23	36	77	77	99	99	100	101	108	108	108
Test	6	0	6	2	4	3	1	1	3	(6)	4	3	(3)	2	2	2	3	5	4	(4)
Retest	0	3	5	2	4	2	1	2	6	4	5	3	3	2	2	2	3	5	4	2

The dowsing sucess due to stable skills? The Scheunen experimenters did not investigate this question specifically, only 9 of 43 subjects participating more than once (considering here only full sessions comprising 10 trials). Four of them participated twice, 5 participated 3, 4, 5, 6, and 11 times, respectively. Nevertheless, the resulting hits per session obtained by the nine dowsers, based on our standard definition, may be subjected to test-retest correlation. Table 1 shows the data. A nonparametric retest-correlation yields tau = 0.42, $z = 2.23$. The result is very significant ($p = 0.01$).

The stability of dowsing performance was also tested *within* sessions. A total of 58 full sessions (consisting of 10 trials each) yields 58 pairs of hit scores, summed over the first 5 and the second 5 trials, respectively. A split half tau is 0.18, $z = 1.6$, $p = 0.05$. An odd-even number division of trials within sessions yields: tau = 0.24, $z = 2.2$, and $p = 0.01$. Hence, within sessions, the dowsers' successes and failures were also significantly consistent ("reliable"). These results contradict Enright's conclusion based on his eyesight: "Reproducibility by a given individual seems to be absolutely lacking" (p. 366).

Our reanalysis of König and Betz's Scheunen data have revealed more dowsing skill than that suggested by Betz's original analysis. Further, the dowsers' skill was shown to remain relatively constant within and across experimental periods. Finally, it seems that the dowsers identified a virtual pipe at mirrored distances from the real hidden pipe and that these distances are correlated with some aspect of the barn geometry which thus confused the dowsers. Open field experiments would probably have yielded larger dowsing successes. On the other hand, the deflection of signals as indicated in this study suggests systematically varying the pertinent conditions in future dowsing experiments. The physical mechanisms underlying positive dowsing results as observed in the barn might thus come into focus.[7]

1. König, H. L., Betz, H.-D.: Der Wünschelruten-Report. Wissenschaftlicher Untersuchungsbericht. München: Eigenverlag 1989
2. Enright, J. T.: Naturwissenschaften *82*, 360 (1994)
3. Hays, W. L.: Statistics, p. 456. London: Holt 1972
4. Wortmann, C. B., Rabinowitz, in: Evaluation Studies Review Annual (Sechrest, ed.). Beverly Hills: Sage 1979
5. Cohen, J.: Statistical Power Analysis from the Behavioral Sciences. New York: Academic Press 1970

[7] Dr. Joop M. Hootkopper gave helpful comments on an earlier version of this paper.

BUCHBESPRECHUNGEN

Fraktale, Chaos und Selbstähnlichkeit. Von M. Schroeder. Heidelberg: Spektrum 1995. 439 S., DM 78,–.

Das umfangreiche Inhaltsverzeichnis am Anfang des Buches vermittelt einen aussagekräftigen Ausblick auf die zahlreichen Gebiete der Wissenschaft, die man beim Lesen durchstreifen soll. Selbst wenn man daraufhin die Lektüre mit hohen Erwartungen beginnt, ist kaum mit einer Enttäuschung zu rechnen: Die Vielfalt der behandelten Themen und aufgezeigten Querverbindungen zwischen unterschiedlichen Zweigen der Wissenschaft ist schwer zu überbieten, und von den leicht verständlichen, originell verfaßten Texten ist man schnell vereinnahmt.

Im Vorwort wird die im Umfeld von Chaos und Fraktalen allgegenwärtige Skaleninvarianz als eine Form der Symmetrie eingeführt und mit anderen Symmetrien verglichen, die in den Naturwissenschaften von Bedeutung sind. Durch eine geschickte Auswahl von Beispielen stellt der Autor im einleitenden Kapitel dar, daß Phänomene der Invarianz zu allen Zeiten in den Naturwissenschaften beobachtet und mathematisch untersucht wurden. So wird anhand eines einfachen Beweises zum Satz von Pythagoras, den Einstein im Alter von elf Jahren formuliert hat, die Bedeutung von Ähnlichkeit und Skalierung veranschaulicht. Dieses und andere Beispiele lassen den Leser erkennen, daß die Grundideen der fraktalen Geometrie eine weitreichende Tradition haben. Ihre Methoden, vor allem ihre unterschiedlichen Dimensionsbegriffe, ermöglichen ein tieferes Verständnis vieler klassischer Fragestellungen.

Zusätzlich zum Kanon der Themen, die üblicherweise mit Chaos und Fraktalen in Verbindung gebracht werden und zu denen etwa die Iteration von Funktionen, Julia-Mengen, die Mandelbrot-Menge, Periodenverdopplung und viele weitere Themen gehören, behandelt der Autor zahlreiche andere Erscheinungsformen der Selbstähnlichkeit in Physik, Mathematik und Statistik. Bemerkenswert sind aber vor allem die vielen einfachen Gedankenexperimente, die Situationen aus dem Alltag aufgreifen und denen jeder Leser ungeachtet seiner naturwissenschaftlichen Vorbildung folgen kann. Zum Beispiel wird der Frage nachgegangen, wie sich der Schaltrhythmus der Ampeln in einer