Water Dowsing: the Scheunen Experiments

J.T. Enright
Neurobiology Unit 0202, Scripps Institution of Oceanography, La Jolla, CA 92039, USA

A massive experimental program was undertaken in 1987 and 1988, to investigate whether water dowsers can, as they claim, detect water from a distance by extraordinary means. Funded by the German government (BMFT) for DM 400,000, the carefully planned research was conducted by university scientists from Munich, and involved nearly 10,000 tests using about 500 dowsers. The results of that study have been interpreted as demonstrating that a real dowser phenomenon exists— that certain individuals achieved remarkable success rates, apparently based on processes that present-day science cannot explain: an interpretation that has been widely publicized. A reexamination of the data on which that conclusion was based, however, indicates that no persuasive evidence was obtained for a genuine, reproducible dowsing skill. The absence of reproducibility suggests that the success rate cannot reasonably be attributed to chance.

Public attitudes in North America toward water dowsing (also known as water witching, or divining) involve a remarkable ambivalence. As is true of the rest of the Western world, a majority of the water wells that are drilled in the USA involve consultation with a dowser, and many such "experts" are available: in 1967, for example, the number of practicing dowsers in the USA was estimated to be about 25000 [1]. Nevertheless, most of the educated American public, serious scientists as well as nonscientists, would probably characterize water dowsing as a bold over from medieval superstitions: nothing more than an unreliable folklore. The usual justification for such skepticism is that no plausible physical explanation has ever been offered for the stimuli to which a dowser, with his "divining rod", might be responding. When considered objectively, however, a rejection of dowsing simply because physics and physiology cannot provide an adequate mechanism to account for the phenomenon can be interpreted as scientific arrogance. As open-minded counterargument is that the tradition and folklore of dowsing are not based on its theoretical underpinnings but on the claimed successes of its practitioners; and if the method "works" even though current science cannot explain it, so much the more for science!

But are water dowsers truly more successful than can be accounted for by pure chance? Anecdotal reports of positive results can, of course, be found in abundance, but a survey by the US Department of the Interior of more than 500 publications on dowsing led to the following assessment [2]:

"It is doubtful whether so much investigation and discussion have been devoted to water dowsing in the USA because it is a good technique; rather, it is puzzling because it has failed to produce results that can be accounted for by chance. Dowsing is not a well-defined phenomenon, and it is difficult to see how for practical purposes the entire matter could be more thoroughly dissected, and it should be recognized that further tests by the United States Geological Survey of this so-called "witching" for water, oil or other minerals would be a misuse of public funds."

thing truly extraordinary was involved in the perfor-
mances of at least some of the few water dowsers in-
vestigated. A conclusion of this sort has a significance
that should not be ignored by the broader scientific
community, as it has major reaching implications, not
doing for dowsing itself, but for both physics and physi-
ology. (What form of energy might arise from flowing
water that could serve as an adequate generalized tool
for predicting not only water, but for physics and physi-
ology?)

The overall research program under the Dowser's and Earthwork grants included several different approaches. In one of these, the ability of dowsers to detect artifi-
cial magnetic fields was investigated, with completely
negative results. A deeper analysis of these results
instructs that magnetism is involved in their skills. An-
other set of tests (Lanzetti) evaluated whether there
was any agreement among dowsers in their tendency independently to choose identical loca-
tions along preselected outdoor test paths, presumably
as a response to some unknown local stimulus, per-
haps related to earth's magnetic field. Evidence for nonrandom agreement among dowsers was reported, but this sort of testing is very difficult to interpret, since no correla-
tion with underground water supplies was investigated, and the actual location and nature of the relevant stimuli were completely unknown to the experimenters. This experiment is somewhat comparable with asking a group of people each to choose ten numbers between 1 and 100. If one then finds that their selections show nonrandom agreement with each other, no sensi-
tible person would insist that mysterious stimuli origin-
ating from the preferred numbers themselves (Zah-
tenbrink's) were involved. The third component of the
research, however, the so-called Scheneun experi-
ments, seems to get to the heart of the matter; these
tests were designed to determine whether experienced
dowers can do exactly what they claim: localize the
presence of water from a distance, in the absence of or-
dinary clues. Flowing water was actually present at a
nearby well, unknown to the experimenters, and that portion of the research program seems to have been well designed for its purpose.

A 10-m-long stimulus line was established on the
ground floor of a two-story building, and a short wa-
ter pipe, oriented perpendicular to that line and con-
tained in a tank of water, was far enough from the pipe
was moved back and forth along the stimulus line, so that
for each test the pipe could be repositioned to a loca-
tion determined by the random generation of a comput-
er. In each test, the water pipe was passed through the
pipe at rates of up to 101 min⁻¹, and a geologist might
well ask why flowing water was used, rather than a pipe
of standing water. The use of flowing water is based on the
assumption that under ground water is usually found in "water
dropping," namely, as small wet and porous sediment, as most geologists today believe, as
within the dowser tradition, the stimuli to which
a dowser and his divining rod respond are often thought
to be intimately related to the movement of the wa-
to the water heater. (The widespread existence of
constructed, flowing underground streams of water is
of course, central to the dowser tradition, presumably
making it essential to excavate in almost exactly the
right place; if, instead, one assumes that water may
well be present in extensive, distributed deposits, the
services of a skillful dowser would be irrelevant.)

On the second line, another 10-m-long stimulus line was established, directly above the groundfloor
stimulus line, with measurement markers correspond-
ing to those below; and it was the task of the exper-

imenters to use whatever tools of the trade he preferred, to
indicate in each trial exactly where along the upstream wa-
line he "felt" the pipe to be. The standard protocol
typically involved a series of ten tests (sometimes for-
er), which were usually completed within about 1 h, w-
that the dowser had roughly 5 min available to wander
back and forth along the test line before making each
decision. Tests, while the pipe on the ground floor
was being moved, the dowser was taken to an ad-
jaent upstairs room or the outside building, so as to
minimize any opportunity of hearing noises or wa-
terswelling; thus, it was the dowser who was the
new pipe location
was; and, as a further precaution, a constant in-
terval between tests for moving the pipe was main-
tained, so that no conclusions could be drawn as to
whether the pipe had been moved or not.

Another extremely important element in all the repor-
ted tests is that they were conducted "double-blind,
meaning that neither the dowser nor an experimenter
were in the same room to supervise his activities, knew
the actual location of the pipe below might offer a hint of the answer. During these "critical" test series, the dowser
was not informed whether the pipe was in fact present or not. If the pipe was located correctly, the dowser
was permitted to terminate a session if he
became tired or felt they had lost their ability to com-
transport, and the published data make it appear that the
mentor's water pipe. There is, of course, residual
and legitimate concern about whether some of the
dowser's claims have been the result of the encour-
ment or procedure design; in this kind of project, the tem-
pronunciation to fraud cannot be overlooked. Particularly
when even dramatic performances were purportedly ob-
tained from only a few individuals, such concern
merely difficult for physical scientists to anticipate all
the subtle ways in which they might be deceived. As a
sentence precaution against this possibility, the investi-
gators invited a professional magician — one of those
in a better sense of deceit — to examine the experimental setup
for possible weaknesses.

Nevertheless, certain subtle variables were incomplete-
ly controlled during the experimentation, the most ob-
vious of which is the sound associated with the circu-
lating water below. In at least some cases, the water pipe
was intentionally made turbulent, so that if one had
searched along the upstream test line with a very sensi-
tive parabolic microphone, which monitored a broad
frequencies band of sound, it seems likely that such
local differences would have been detectable in parts of the test.

A First Look at the Reported Data

The experimental results from the Scheneun experi-
ments were tabulated in an extensive appendix to the
Final Report (3), pp. 89—101). The outcomes of more
than 800 critical tests, all conducted in the same test
series, are presented, including data from 43 of the 50 preselected
dowers, in a total of 104 test series. For each test se-
ries, the participant is identified by number, and the
actual location of the pipe (typically 100 ft below
the surface) is given together with the corresponding primary (and
sometimes secondary) locations chosen by the dowser.

Location data are given to the nearest decimeter and
100 possible values along each of the two 10-m lines.
Even a first glance at these numbers provides clear
support for the first of the general conclusions cited
above: that overall, the dowsers did very poorly in
a "real core" of dowsing phenomena has been proven. That kind of question can, instead, be more rigorously examined by statistical analysis.

**Conventional Statistical Analysis**

The purpose of most statistical analysis is to distinguish between properties of a data set that could well have arisen by random chance alone, and those that probably represent "real" phenomena. By definition, "real" or "statistically significant" effects are those that are not easily reproducible; if the experiment were to be carefully replicated elsewhere by other investigators. It is important to realize that, at the outset, there are dozens of different ways that a body of data like those from the Scheunen experiments might be analyzed statistically, any one of which would be fundamentally correct and appropriate; and in general, these different ways of examining the data can be expected to give somewhat different answers. It would be an abuse of inferential statistical analysis, however, to search through the armory of statistical tests available until one finds a test that, when applied to the available data, gives the answer one is hoping for; instead, rigorous inferential analysis requires that the test to be relied upon has been chosen in advance, before any of the data have been examined. That demand arises because, with a sufficiently extensive search among analytical techniques, almost any set of unrelated numbers will lead to a purported "statistically significant" outcome that is spurious. (In "exploratory" data analysis [4-6], on the other hand, it is entirely appropriate to examine a data set in many different ways at one chooses. Then, however, levels of "statistical significance" lose all objective meaning.)

The statistical procedure used in the Final Report for the Scheunen experiments is a special, unconventional and customized analysis[1], and the report does not in any explicit way state whether this choice of statistical procedures was made before any of the critical experiments was performed. Because of this concern, I have undertaken a series of other, more ordered analyses to determine whether the conclusions of the researchers will withstand a rigorous scrutiny.

Along the alternative that I have considered are calculation of the correlation coefficients, as well as the fit of regression lines, for data like those in Fig. 1; analyses based on the binomial distribution (with two independent variables) for the critical criterion for "success": a choice within ±1 m pipe location, or a choice within ±0.5 m; chi-square procedures; and Kolmogorov-Smirnov tests. Since such analyses were undertaken, based on all the available data, the results were uniformly discouraging the dowsing hypothesis. For example, the product-correlation between observed and expected frequency for the "best-dowsers" data in Fig. 1, is only about 0.1 — a value that cannot be confidently distin-

1 In the Final Report [3], the statistical significance of a given test series for a given dowser was calculated as follows: with a set of n tests [5 n (n-1)] resulting in m values for the distance, D, between pipe location and chosen location for each test test a set of s, based on the following criteria and categories:

- If [D] < 0.625, then [x] = 0.625.
- If 0.625 < D < 0.642, then [x] = 0.658.
- If 0.642 < D < 1.015, then [x] = 0.8154.
- If 1.015 < D < 1.500, then [x] = 0.925.
- If D > 1.500, then [x] = 0.900.

The values 0.625, 0.658, 0.8154, and 0.925 were chosen because they are the probabilities of each single value, determined as the ratio of its width (0.4286 or 0.8572) to the sum of the test series with a correlation coefficient was less than 0.1 in 20% of the end of the test line, correction for "end effects" must be incorporated in calculating the actual significance of the test.

Statistical Analyses in the Final Report

The Final Report nowhere specifies which of the 43 dowzers should be regarded as the "eigen" or "best" dowzer, and who constitute the "real core" of the dowsing phenomenon, but a reasonable criterion is available by which to judge this. Each test series was analyzed separately (see footnote), and Table 6 of the Final Report summarizes those 104 calculations for the standard barn experiments with a derived "probability" for each test series, along with identity of the dowzer. In that table, three of the test series, each from a different dowzer, were assigned probabilities of less than 0.01; and another four test series, from three other dowzers, were assigned probabilities between 0.01 and 0.03. For purposes of further examining the results, it seems reasonable to assume that the "best" dowzers are the three that achieved the three most significant test series (Dowzers # 18, # 99, and # 108); and that the four next-best results (with assigned probabilities between 0.01 and 0.03) came from the three "second-best" dowzers (# 23, # 89, and # 110).

The relationship between pipe locations and the dowzer's choices from the three "best" test series are illustrated in Fig. 2A. It is evident there that, although there were many conspicuous errors, some 14 of the 26 positions chosen agreed with the location of the pipe to within 1 m, and 11 of the 26 agreed to within 0.5 m. On the basis of this kind of agreement, and the statistical analysis of the Final Report, the three people who participated in those series can reasonably be regarded as the "creme de la creme" of dowzers, the three very best of a highly select group: part of the roster of the dowzer phenomenon referred to in the Final Report. The results shown in Fig. 2A do indeed look impressively favorable for the capabilities of these particular dowzers; and in the Final Report, the only graphical presentation of results from the Scheunen experiments (3), Fig. 18) is a plot of those data in Fig. 2A that are shown by filled circles. The novel statistical test utilized in the Final Report classifies the results from each of these three test series as having a probability of less than 0.01 of being due to chance alone. Despite these calculations (and the general impression given by the conspicuousness of the errors evident in the plot of 2A), a skeptic might doubt that the correlation coefficient between observed and expected, even for the highly selected data in Fig. 2A, is sufficiently "statistically significant" at even the 10% level. An advocate of the dowsing hypothesis might counter, however, that the correlation coefficient signifies a potentially real effect in the population from which the test series was drawn. Overall, the fact that many of the choices in these test series were remarkably close to
the actual pipe location; that, for example, in two of these test series, the dowser chose a location within 50 cm of the pipe's location four times out of ten. The skeptic might then reply that, according to the binomial expansion (defining "success" as a choice within ±50 cm of correct), it is not at all surprising (p > 0.50) to find two such cases (four successes in ten attempts) among the 104 test series reported. Such differences of opinion illustrate the ambiguity that arises if one sets aside the requirement, for rigorous statistical inference, of preselecting the (single) test to be relied upon. In any case, a major problem arises if one wants to interpret the results from these three test series in terms of the special skills of these particular dowsers: the fact is that these same three individuals participated in several other experimental series at other times, and their performances in those other tests were by no means as impressive as those in Fig. 2A. The rest of their results (same dowsers as in Fig. 2A) are shown in a composite plot (Fig. 2B); and the results from all tests in which each of these three dowsers participated are summarized in Fig. 2, parts C, D, and E. It is difficult to avoid the impression from these graphs that, overall there was little if any relationship between the locations chosen and the location of the pipe. (The correlation coefficient for Dowser #99 is +0.06, and those for Dowser #18 and #110 are slightly negative.) None of these values would be considered "statistically significant": p < 0.50, but of course, such a statement about "probability" should be discounted, being based on a-posteriori testing of the data.) 

In any case, it seems quite clear that a dowser did unusually well on one occasion was not particularly likely to do well in another comparable test series, on another occasion. The scatter in these results demonstrates that it was not three particular dowers ("some few people") who consistently did well in locating the pipe, but instead, that within the array of 104 test series available, one can find three in which many of the choices were relatively close to the pipe location. Reproducibility by a given individual seems to be accurately lacking. What about the overall success rate of the next-best group of dowers: the three individuals who in single-test series achieved results that were assigned "probabilities" between 0.01 and 0.03 (Dowser #23, #85, and #110)? The results from their best performances (two test series by Dowser #23) are illustrated in Fig. 3A, and again (as in the case of Fig. 2A), these "best" results, when viewed by themselves, look impressive. When such results are placed in the context of what these same individuals did in other test series, however, the evidence in favor of significant abilities blends into a cloud of scattered and seemingly random choices (composite results in all their other tests shown in Fig. 3B, and overall data from each of these three dowers in Fig. 3, parts C, D, and E). From these data, it appears that the three "second-best" dowers were neither appreciably better — nor appreciably worse — than the three best.

---

**Fig. 2.** A) Dowers' choices from the three "best" test series (those with strongest relationship to location of the pipe, assigned probabilities of <0.01 in Final Report); B) Dowers' choices from the three "second-best" test series by the same three dowers whose "best" data are plotted in (A): C), D, E) Data from (A) and (B) combined, and segregated by dowser; ♦ = Dowser #99; ▲ = Dowser #85; △ = Dowser #110. Data from (C), (D), and (E) represents double choices, from tests in which two locations were selected as being equally likely.

---

**Fig. 3.** A) Choices made in the four "second-best" test series (assigned probabilities between 0.01 and 0.05 in Final Report); B) Dowser #23 (two test series); C) Dowser #110; D) Dowser #85. Data from (A) and (B) combined, and segregated by Dowser; ♦ = Dowser #99; ▲ = Dowser #85; △ = Dowser #110. Data from (A) and (B) combined, and segregated by dowser; ♦ = Dowser #99; ▲ = Dowser #85; △ = Dowser #110.

---

**Fig. 4.** Lower three rows of data points show magnitude of errors, relative to the pipe, made by the three "best" dowers (left panel) and the three "second-best" dowers (right panel). Upper three rows of data show how far these pipe locations were from the middle of test line in these same tests; hence, they illustrate errors that would have been made by the dowers, had they always chosen the middle of test line as their "guess."
Discussion

The experiments described here represent the most extensive and carefully controlled studies ever undertaken to investigate the capability of dowsers to detect water at nearby, established locations. If water dowers—even some small fraction of them—possessed the ability implied by the results, they should have had a very good chance to demonstrate it. The pipe to be detected was only 3 or 4 m from the dowser, rather than tens to hundreds of meters belowground, and the group involved a simpler assignment than the kinds of field problems that water dowers regularly confront. Those who were actually tested in the final, critical experiments were selected from a much larger pool of candidates on the basis of what was judged to be good performance on preliminary trials, so these should have been the best of available experts. Furthermore, each dowser was permitted modest variations in testing conditions (e.g., velocity of flow, and nature of the fluid in the pipe), which conformed with those situations in which he had done well in the preliminary trials—a measure that appears favorable to a successful outcome of the testing. One of the common objections to scientific tests of unusual sensory capacities or extraordinary phenomena based on paranormal abilities (ESP) is that the presence of a hostile audience can make it difficult for specially gifted people to perform at their best; but the more published description of this experimental study, as well as the conclusions to which the investigators themselves came, indicate that this project was conducted in an atmosphere that was anything but hostile to the claims being tested.

This study also had many features that should please the skeptic. A variety of proper precautions were taken to assure objectivity of the testing procedure; mechanically randomized locations of the target; "doubleblind" arrangements, so as to avoid subtle, unintended signals between experimenter and dowser; no feedback to the dowser about the quality of his performance during testing; isolation of the dowser between tests, to eliminate several possible sources of unintended transfer of information about the target location; and, at all times, measures against sound experimental design, and involve the kinds of precaution that a thoughtful skeptic should expect to see— or at least hope for—when someone is testing a controversial hypothesis that challenges established scientific principles.

Had the outcome of such a large, well-planned study been unequivocally positive, it had demonstrated strength (and explainable skills) in most of the dowers, the outcome should be expected to serve as a springboard for intensive follow-up research by physicists as well as physiologists, to explore what mechanisms might be responsible for these capabilities. Had this kind of data been obtained, then someone who remains a skeptic must be forced to invoke some undetected cheating by the dowers to support his position. That sort of success was not, however, achieved as the Final Report recognizes, the vast majority of the participants who enrolled in the critical experiments (who had not been preselected as the best 10% of a much larger group of people, all of whom thought that they had the ability to detect and localize hidden water supplies) did very poorly; most of their performances in the critical trials were so poor that they could not be distinguished from the results of random chance.

Nevertheless, the researchers who conducted the study were persuaded by their data and their analyses that they had uncovered a small but "real core" of the water-doweling phenomenon— that some few individuals showed an extraordinarily high success rate. The examination of those data described here indicates that this conclusion rests on very flimsy grounds indeed. In the Final Report, attention was focused on one small number of unusually good performances (Fig. 2A and 3A). If one relies only on probabilities from the unusual, customized statistical test of the Final Report, then it is relatively unlikely that a small subset of the results could have arisen due to chance. Other, more standard ways of examining the overall data, however, support the opposite interpretation: that obtaining a few such results among the 104 test series is not at all surprising. Thus, the interpretation of those few exceptional test series ("unlikely" or "very likely" to be due to chance) depends more on choice of statistical procedures (and when that choice was made) than on the data themselves; and there is no objective way of deciding, after the fact, which interpretation is more credible.

If doweling is a "real" phenomenon, however, the most important, central expectation is that, in some way success must be reproducible — and the overall results certainly do not meet that criterion. As shown in Figs. 2A and 3A, even the most "successful" test series were obtained by people who could not themselves replicate that sort of success. Overall, those individuals who were successful at all were not any better than those of the rest of the dowers, no better than one might expect due to chance alone. Perhaps the most interesting conclusion that can be drawn is that even the dowers who, on single occasions, managed to perform at considerably better-than-chance levels, could on average have done no better. What they did, in fact, if it had actually chosen the midpoint of the test line is end and every test. It is my impression that the Schoenex experiments have devastating implications for the art and profession of water dowing. Although the research was conducted by a group fully sympathetic to the cause of truth and freedom of inquiry, and has been interpreted by them as indicating that successful doweling is indeed a real phenomenon, even the most favorable interpretation of the experiments is discouraging. The data indicate that if it is appreciably better than chance levels are to be obtained, one must engage a very select dowel (one with at least 99.9% of his competitors) on one of the very good days (and he cannot tell you whether it is a good day or not); and even then, that super-expert, a best day, is apt to be badly wrong about half the time locating a water source that is only 3 or 4 m away.

Would you yourself be willing to pay someone this advice, if he attempted to demonstrate his command by showing a graph resembling one of those included in parts C, D, and E of Fig. 2 and 3? The record shown in Fig. 4 could be interpreted as suggesting in the absence of other information, it would be better strategy for someone who is planning to drill oil, to sink his hole right in the middle of the area under test, rather than to rely on advice from a dowel to make the dowel claim to have been on certain past occasions. This interpretation depends, of course, on the assumption that there single best place to drill, and that the closer one that location, the better the expected outcome.

Conclusions

The Schoenex experiments involved such a no-scale test program, which incorporated both rigorous allowances favoring the doweling hypothesis (a careful, rigorous experimental program, a definitive answer is finally available to the central, age-old questions about water doweling. Briefly stated, the conclusion is that even with very extensive testing, by others sympathetic to the cause, no persuasive evidence could be found for reproducibility of the "water doweling" phenomena, neither individual reproducibility nor intraindividual reproducibility. Instead, as has now been demonstrated that the ability to locate water from a distance by extraordinary stimuli exists, that skill cannot be reproducibly demonstrated across a select group of 6 experts from among 500 dowers who all think that they have the ability (a conclusion consistent with the summary in the Final Report).

2. In those few cases in which a single series of tests suggests that a given dowler may perhaps have better-than-chance abilities, similarly good results are not reproducible by that same individual in further comparable test sessions (a conclusion which contradicts the summary of the Final Report).

Properly considered, then, these are answers as definitive as experimentation could ever provide. Reproducibility lies at the core of successful experimental science; and if a phenomenon is not reproducible, even for select individuals, what possible gain could come from further, similar experiments, no matter how expensive the program? Thus, the Schoenex experiments are not only the most extensive and careful scientific study of the doweling problem ever attempted, but--if reason prevails—they probably also represent the last major study of this sort that will ever be undertaken.

This does not, of course, constitute a rigorous refutation of the doweling hypothesis. A universal negative can never be proven by observation, and it remains conceivable that individuals exist who can indeed reproducibly detect water from a distance by extraordinary means. If so, however, one must assume that they are so rare that none turned up in the sample of 500 candidates, all of whom thought that they had the required ability. Hence, if unanswerable, it still leaves open the question of whether distinguishing them from the unskilled appears to be a hopeless task. This leads to a valuable insight: whether one prefers the interpretation that truly skilled dowers exist, who are so rare that none was found in the Schoenex experiments, or whether one instead prefers the interpretation that the ability claimed by dowers does not exist, is no longer a question of evidence. The choice is simply a matter of arbitrarily deciding between hypotheses that have identical consequences.

Research supported by Grant IBN 93-1303 from the National Science Foundation. Valuable comments on the manuscript were made by Dr. J. W. Westab.}