NATUR
WISSENSCHAFTEN AUFSÄTZE

Water Dowsing: the *Scheunen* Experiments

J. T. Enright

Neurobiology Unit 0202, Scripps Institution of Oceanography, La Jolla, CA 92093, 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 400000, the carefully planned research was conducted by university scientists from Munich, and involved nearly 10000 tests, using about 500 dowsers. The results of that study have been interpreted as demonstrating that a real dowsing 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 entire research outcome can 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 holdover from medieval superstitions: nothing more than 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. An 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 worse 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 bestowed on any other subject with such absolute lack of positive results. It is difficult to see how for practical purposes the entire matter could be more thoroughly discredited, and it should be obvious to everyone that further tests by the United States Geological Survey on this so-called "witching" for water, oil or other minerals would be a misuse of public funds".

In Germany, where water dowsing apparently originated some time in the 16th century, a divining rod – the tool of the trade – is called a Wünschelrute (literally wishing rod), a term which seems to imply a similarly skeptical attitude. Nevertheless, a considerably more tolerant opinion can often be encountered in Germany than in America among both professional scientists and educated nonscientists: not that dowsing is widely and confidently accepted as a proven phenomenon, but it seems to be tolerated as an expression of the attitude: "There are more things in heaven and earth..." than modern science can fully account for; perhaps dowsers actually are able to respond to some stimulus not yet understood by modern science. In Germany, there is widespread interest among laymen in what dowsers refer to as *Erdstrahlen* – some sort of radiant energy purportedly associated with underground water (and other minerals as well), to which dowsers are thought to respond. Although the exact physical nature of these rays remains unspecified, their presumed existence is regarded by some as a major gap in current scientific knowledge.

A Scientific Study of Dowsing

In view of this widespread cultural acceptance (or at least tolerance) of water dowsing in Germany, it should not be surprising that in 1986 the BMFT (Bundesministerium für Forschung und Technologie) provided DM 400000 to several established scientists for an experimental investigation of the phenomenon. The funding was one element of a much larger research program entitled Unconventional Methods of Cancer Control, a connection based on the claim by some radiesthetists (as European dowsers sometimes call themselves) that their services can help protect homeowners from the adverse effects of *Erdstrahlen* on health. Two of the principal investigators on that grant for research on Earthrays and Dowsers were physicists, one from the University of Munich, one from the Technical University of Munich, the third being a professor of pharmaceutical biology at the University of Munich. The objective of their study was to determine whether careful scientific research on water dowsing could lead to convincing evidence for or against the existence of real, reproducible effects: an experimental examination of whether dowsers really can somehow detect water with a success rate greater than can be accounted for by chance alone. It was not the intent of the research, even in the event of a favorable conclusion, to provide a theoretical explanation for such effects, and this seems to be an eminently reasonable approach. There is no point in seeking a physical explanation for earthrays and their possible connection with cancer until persuasive experimental evidence is available to demonstrate that dowsers really can do as they claim, since it is those claims that gave rise to the hypothesis of earthrays in the first place.

This research project was no doubt the largest carefully controlled scientific study of dowsing ever conducted. Some 500 candidates who claimed skill as dowsers were investigated in some 10000 individual tests. Much of that effort was devoted to preliminary experiments, which were intended to develop the most sensitive methods available and to select the best candidates for the final stage of the project: "critical" tests with rigorous methods and extremely careful experimental precautions. The most interesting and enlightening of these final experiments, which are to be examined in more detail here, were referred to as the *Scheunen* (barn) Experiments; as the name suggests, these tests were conducted in a two-story building that had previously served as a barn. In these tests, the dowsers were required to select a location, on the second floor of the building, that they thought was directly above a water pipe located on the ground floor. That pipe could be moved back and forth across the floor on the ground level, to a location that was chosen randomly for each test. (More experimental details are given below.)

The Final Report on this research project was submitted to the granting agency (BMFT) in 1990, with a title that almost seems intended to conceal rather than disclose its content [3]: „Setting up and operation of test arrays with artificial variable low-energy fields for the study of the response in biological macrosystems." By use of italics, the summary of the report emphasized the two primary conclusions, one negative, the other positive:

"I: The success rate of average dowsers in the tests conducted was poor and in most cases indistinguishable (or nearly so) from chance;

II: Some few dowsers, in particular tasks, showed an extraordinarily high rate of success, which can scarcely, if at all, be explained as due to chance." ([3], p. 5)

Elsewhere on the same page, the second, positive conclusion was elaborated: "... in every sort of test conducted, there were some few people who showed location-dependent responses, some with good and some with extraordinarily good reproducibility, which, in their departure from chance expectations, were highly significant."

As a final rephrasing of that second, positive, and widely publicized conclusion, the summary says:

"... a real core of dowser-phenomena can be regarded as empirically proven. ..." ([3], p. 5)

As these quotations demonstrate, the investigators had become thoroughly convinced by their results that dowsing is a reproducible phenomenon; that some-

Naturwissenschaften 82, 360–369 (1995)

thing truly extraordinary was involved in the performances of at least some few of the water dowsers investigated. A conclusion of this sort has a significance that should not be ignored by the broader scientific community; there are far-reaching implications, not just for dowsing itself, but for both physics and physiology. (What form of energy might arise from flowing water that could serve as an adequate stimulus for precise determination, from a distance, of the direction to its source? What sort of sensory system might be able to respond to such stimuli, which would presumably be exceedingly weak? How can a high-tech instrument be built to amplify these stimuli?) Fortunately, the Final Report also includes extensive appendices in which the experimental observations are summarized, and this permits others to decide for themselves whether such revolutionary conclusions are fully warranted by the data.

The Scheunen Experiments

The overall research program under the *Dowsers and Earthrays* grant involved several different approaches. In one of these, the ability of dowsers to detect artificial magnetic fields was investigated, with completely negative results – but water dowsers do not necessarily insist that magnetism is involved in their skills. Another set of tests (*Laufbrett* experiments) evaluated whether there was any agreement among dowsers in their tendency independently to choose identical locations along preselected outdoor test paths, presumably as a response to some unknown local stimulus, perhaps related to earthrays. Evidence for nonrandom agreement among dowsers was reported, but this sort of testing is very difficult to interpret, since no correlation 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 sensible person would insist that mysterious stimuli originating from the preferred numbers themselves (*Zahlenstrahlen*?) were involved. The third component of the research, however, the so-called Scheunen experiments, seems to get to the heart of the matter; these tests were designed to determine whether experienced dowsers can do exactly what they claim: localize the presence of water from a distance, in the absence of ordinary clues. Flowing water was actually present at a nearby hidden location, known only 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 water pipe, oriented peripendicular to that line and connected to hoses, was fastened to a wagon that could be moved back and forth along the stimulus line, so that for each test the pipe could be repositioned to a location determined by the random generator of a computer. In most tests, water was pumped through this pipe at rates of up to 10 l min^{-1}, and a geologist might well ask why flowing water was used, rather than a tub of standing water. The answer lies in the tradition of dowsing; the remarkable assumption is that underground water is usually to be found in "water arteries" (*Wasseradern*) – rather than as extensive pools in porous sediment, as most geologists today believe; and within the dowsing tradition, the stimuli to which a dowser and his divining rod respond are often thought to be intimately related to the movement of the water through the *Wasserader*. (The widespread existence of constricted, flowing underground streams of water is, of course, central to the dowsing 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 become irrelevant.) On the second floor of the barn, another 10-m-long line was established, directly above the groundfloor stimulus line, with measurement markers corresponding to those below; and it was the task of the dowser, using whatever tools of the trade he preferred, to indicate in each trial exactly where along the upstairs test line he "felt" the pipe to be. The standard protocol typically involved a series of ten tests (sometimes fewer), which were usually completed within about 1 h, so that the dowser had roughly 5 min available to wander back and forth along the test line before making each decision. Between tests, while the pipe on the ground floor was being moved, the dowser was taken to an adjacent upstairs room or outside the building, so as to minimize any opportunity of hearing noises that might offer a hint about where the new pipe location was; and, as a further precaution, a constant time interval between tests for moving the pipe was maintained, so that no conclusions could be drawn from that interval about how far the pipe had been moved. Another extremely important element in all the reported tests is that they were conducted "double-blind", meaning that neither the dowser nor the two experimenters, who were in the same room to supervise his activities, knew the actual location of the pipe below. During these "critical" test series, the dowsers received no indication about their success or failure; they were, however, permitted to terminate a session if they became tired or felt they had lost their ability to concentrate, and the published data make it appear that this happened relatively often. Some dowsers often noted more than one location along the line where they "felt" something, and in such cases they were encouraged to select only one of the locations as the main choice, with the other choice(s) being recorded but ignored in the final data analyses. In a small fraction of the tests, no preference between two locations could be made.

Several superficial aspects of the test situation were varied from one dowser to the next; this was done intentionally so that each person would be given an optimal opportunity to demonstrate his capabilities. In preliminary experiments, as well as in the "training" tests (in which feedback about success or failure *was* provided), potential variables included the material of which the pipe was constructed, the fluid in the pipe and pumping system (fresh water, salt water, or air), the rate of fluid flow, the extent of turbulence in the fluid flow and, within limits of the available space, the exact location of the test and stimulus lines. Dowsers who were accepted into the final, critical-test program were those who had shown some indication of success in the preliminary tests; and for each accepted candidate, those variables from the preliminary tests (type of pipe, type of fluid, etc.) that had led to the best results were used for subsequent critical tests with that individual.

There can be no doubt that the experimenters took generous precautions to avoid the objection that the research was biased against successful dowsing. "Less skillful" dowsers were eliminated at the outset: after preliminary testing of some 500 available candidates, only 50 dowsers were selected to participate in the final critical experiments. Thus, if the preliminary tests gave any basis for judgment, one should expect the participants to be the most skillful 10% of the candidates. The preliminary experiments also assured that each dowser was familiar with the experimental setup, and would be permitted to deal with the stimulus that had previously proven optimal for him. Having been through preliminary testing, the dowsers' voluntary participation in the final, critical tests can be taken to indicate that they considered the experiments a fair test of their abilities, and were convinced that they could be successful.

On the other hand, the experimental design incorporated a variety of features that are important from the viewpoint of a skeptic. The computer-generated random locations assured that the dowsers could not successfully attempt to outguess the experimenters; the double-blind design avoided the possibility that the dowsers would receive any unintentional hints about where the pipe was located from subtle behavior of the experimenters; and precautions were taken to prevent accidental transfer of information during displacements of the water pipe. There is, of course, residual and legitimate concern about whether some of the dowsers might have been clever enough to defeat the experimental design; in this kind of project, the temptation to defraud cannot be overlooked. Particularly when remarkable performances were purportedly obtained from only a few individuals, such concern grows, because even a few cheaters would be enough to produce a few surprising results. When confronted by a clever, strongly motivated candidate, it might be extremely difficult for physical scientists to anticipate all the subtle ways in which they might be deceived. As a sensible precaution against this possibility, the investigators invited a professional magician – one of those experts of deceit – to examine the experimental setup for possible weaknesses.

Nevertheless, certain subtle variables were incompletely controlled during the experimentation, the most obvious of which is the sound associated with the circulating water below. In at least some cases, the water flow was intentionally made turbulent, so if one had searched along the upstairs test line with a very sensitive parabolic microphone, which monitored a broad frequency band of sound, it seems likely that subtle local differences would have been detectable in some parts of the sound spectrum, which could indicate where the pipe was located. As shown below, however, there is very little reason in the experimental results to suspect that any of the dowsers consistently benefited from this sort of noise. Overall, the experimental design indicates that the investigators did their best to give the dowsers a fair chance, but also that they took many reasonable precautions of the sort appropriate for rigorous testing of the purported extraordinary abilities.

A First Look at the Reported Data

The experimental results from the Scheunen experiments 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 barn, are presented, including data from 43 of the 50 preselected dowsers, in a total of 104 test series. For each test series, the participant is identified by number, and the actual locations of the pipe (typically ten values) are given together with the corresponding primary (and sometimes secondary) locations chosen by the dowser. Location data are given to the nearest decimeter, with 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

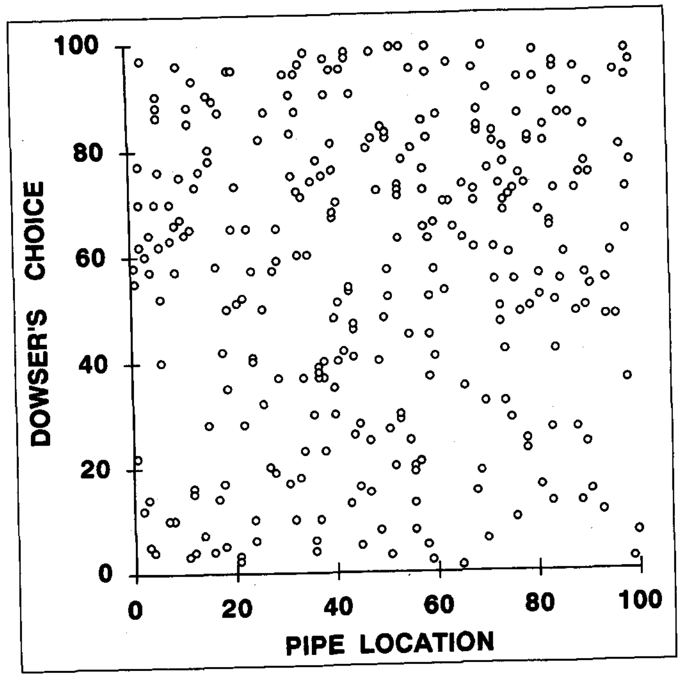

Fig. 1. Relationship between the locations chosen by the six "best" dowsers and the corresponding actual locations of the water pipe. Locations along experimental line (*abscissa*) and test line (*ordinate*) scaled in dm (See text for basis on which "best" dowsers were identified)

matching their choices to the locations of the water pipe.

Some of the observations are presented graphically in Fig. 1, and as can be seen there, the choices were extremely scattered; the results in Fig. 1 were chosen for plotting here because they represent all the data from those six dowsers who can be judged (see below) to be the "best" of the 43. What can one conclude from such a plot? One can easily notice that a majority of the locations chosen are on the upper half of the graph – more than 61% – and such a strong bias is very unlikely to have arisen due to chance, but it is irrelevant to the issue of dowser accuracy. (It is, however, the kind of strong nonrandomness that could contribute to impressive results in tests like those of the *Laufbrett* experiments.) In addition, with a bit of squinting at the graph, one can also notice a weak tendency for the plotted points to aggregate along a 45° diagonal, from lower left to upper right, and this kind of trend indicates a more or less correct match between pipe and chosen location. (This is not surprising, of course, since the data plotted here were selected to include the very best test series.) On the other hand, one might also be able to persuade a willing witness that quite a few points are also distributed along the opposite, downwardly directed diagonal. When data are this scattered relative to expectations, it is not easy to decide, simply from looking at a graph, whether the responses provide persuasive support for the claim that 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 objectively between properties of a data set that could well have arisen due to random chance alone, and those that probably represent "real" phenomena. By definition, "real" or "statistically significant" effects are those that are expected to be *reproducible*, if the experiment were to be carefully replicated elsewhere by other investigators. It is important to realize, at the outset, that 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 purportedly "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 as many different ways as 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-

[1] In the Final Report [3], the statistical significance of a given test series by a given dowser was calculated as follows: with a set of n tests ($5 \leq n \leq 10$), resulting in n values for the distance, D, between actual pipe location and dowser's chosen location, calculate for each test a score, s, based on the following criteria and categories:
if $|D| < 0.2143$ m, $s = 0.7875$;
if $0.2143 < |D| < 0.6429$, $s = 0.558$;
if $0.6429 < |D| < 1.0715$, $s = 0.1854$;
if $1.0715 < |D| < 1.5001$, $s = 0.0295$;
if $|D| > 1.5001$, $s = 0$.
Sum the n values of s, to give a test-series score, S; then determine the percentile ranking of this total score in a cumulative distribution of random expectations based on the multinomial expansion, $(w+x+y+z)^n$, where w, x, y, and z are the probabilities of each single value of s, determined as the ratio of its width (0.4286 or 0.8572 m) to the total length of the test line. Whenever the pipe location was less than 1.5 m from the end of the test line, correction for "end effects" must be incorporated in calculating w, x, y, and z.

ate whether this choice of statistical procedures was ade before any of the critical experiments was perrmed. Because of this concern, I have undertaken a riety of other, more ordinary analyses, to see wheththe conclusions of the researchers will withstand her sorts of scrutiny.

mong the alternatives that I have considered are callation of correlation coefficients, as well as the fitg of regression lines, for data like those in Fig. 1; alyses based on the binomial distribution (with two ernative criteria for "success": a choice within ±1 m pipe location, or a choice within ±0.5 m); chiuare procedures; and Kolmogorov-Smirnov tests. hen such analyses were undertaken, based on all the ailable data, the results were uniformly discouraging the dowsing hypothesis. For example, the productment correlation between observed and expected, en for the "best-dowser" data in Fig. 1, is only about 06 – a value that cannot be confidently distinished (5% level) from zero. This means, then, that one of the standard tests, there is no convincing ination that the pipe location had any relationship th the dowsers' choices.

ch a calculation cannot, however, be taken as a decie answer for two reasons: the data in Fig. 1 have a eater-than-chance likelihood of being judged "sigicant" simply because they are not a random sambut a selected subset (the "best" dowsers, as deterned by another sort of test on the data); and, more portant, my decision to calculate correlation coeffints was made after examining the data, thereby vioing the critical rule mentioned above and potentially sing the outcome by the choice of method. The latof these objections could be raised about any sort etrospective reanalysis of the results. The most that n be concluded from my many analyses is that beuse several ordinary statistical procedures failed to tect unusual consistencies in the dowsers' perforances, the interpretations in the Final Report seem to ve been critically dependent on the selection of a stomized, nonstandard method of statistical analythe investigators would have been led to an oppo negative interpretation if they had instead selected of several more usual methods of data analysis. ence, all claims about statistical significance of the ults are absolutely dependent on the assumption t the choice of statistical method was made *before* data had been obtained from the critical experits (i.e., before 9 April 1987); a later choice of hod would represent "exploratory" data analysis, re probability levels lose their objective meaning.

Statistical Analyses in the Final Report

The Final Report nowhere specifies which of the 43 dowsers should be regarded as the "einige wenige Personen" who were truly skilled at locating water and who constitute the "real core" of the dowsing phenomenon, but a reasonable criterion is available by which to identify them. 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 dowser. In that table, three of the test series, each from a different dowser, were assigned probabilities of less than 0.01; and another four test series, from three other dowsers, were assigned probabilities between 0.01 and 0.03. For purposes of further examining the results, it seems reasonable to assume that the "best" dowsers are the three who achieved the three most significant test series (Dowsers # 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" dowsers (# 23, # 89, and # 110).

The relationship between pipe locations and the dowsers' 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 *crême de la crême* of dowsers, the three very best of a highly select group: part of the *realer Kern* of the dowsing phenomenon referred to in the Final Report.

The results shown in Fig. 2A do indeed look impressively favorable for the capabilities of these particular dowsers; 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 customized 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 of nonrandomness evident in the plot of Fig. 2A), a skeptic might note that the correlation coefficient between observed and expected, even for the highly selected data in Fig. 2A, is only 0.321, which is not "statistically significant" at even the 10% level. An advocate of the dowsing hypothesis might counter, however, that the correlation coefficient penalizes too heavily for occasional gross errors, thereby ignoring the fact that many of the choices in these tests were *remarkably* close to

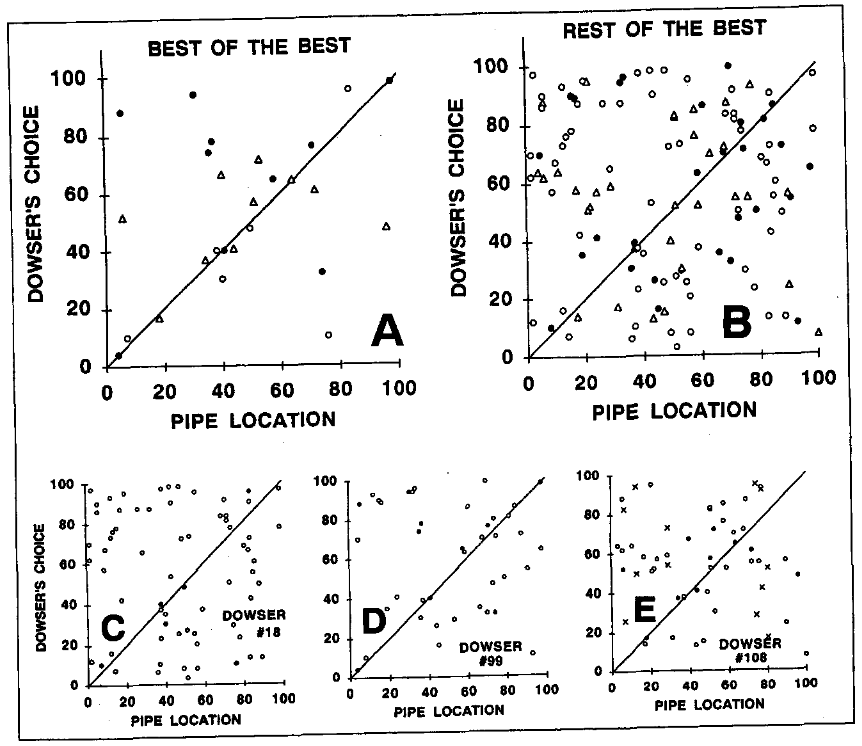

Fig. 2. A) Dowsers' choices from the three "best" test series (those with strongest relationship to location of the pipe, assigned probabilities of <0.01 in Final Report). *Diagonal line* indicates expectation for perfect detection; ○ Dowser # 18; ● Dowser # 99; △ Dowser # 108. B) Choices made in all other test series by the same three dowsers whose "best" data are plotted in (A). C, D, E) Data from (A) and (B) combined, and segregated by dowser; ● choices plotted in (A); *crosses* for Subject 108 represent double choices, from tests in which two locations were selected as being equally likely

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 in 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 Dowsers # 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 who 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 dowsers ("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 acutely lacking.

What about the overall success rate of the next-best group of dowsers: the three individuals who in single-test series achieved results that were assigned "probabilities" between 0.01 and 0.03 (Dowsers # 23, # 89, 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 dowsers in Fig. 3, parts C, D, and E). From these graphs, it appears that the three "second-best" dowsers were neither appreciably better – nor appreciably worse – than the three best.

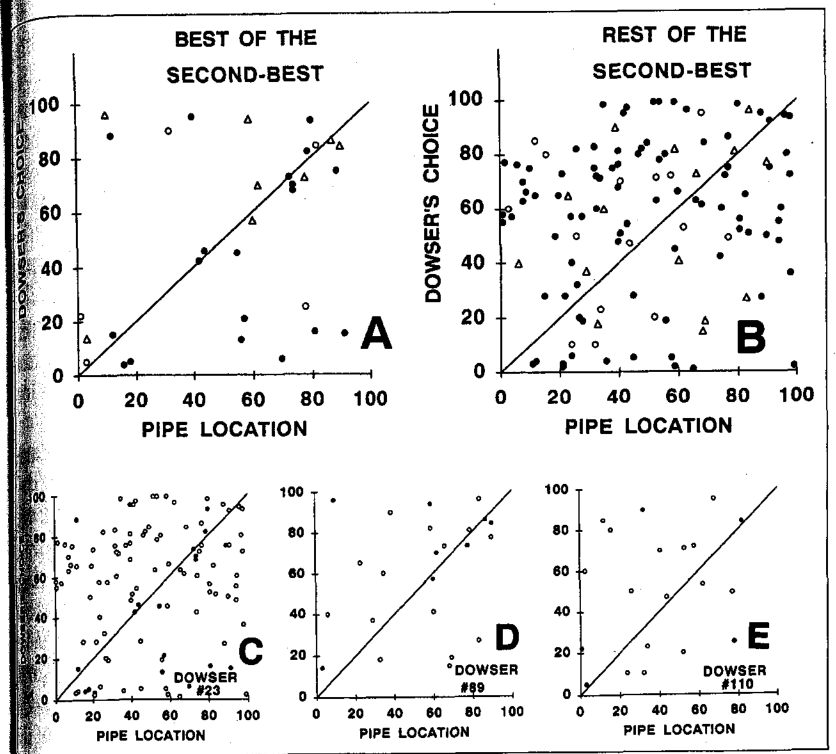

Fig. 3. A) Choices made in the four "second-best" test series (assigned probabilitites between 0.01 and 0.03 in Final Report); ● Dowser # 23 (two test series); ○ Dowser # 110; △, Dowser # 89. B) Choices made in all other test series by the same three dowsers whose data are plotted in (A). C, D, E) Data from (A) and (B) combined and segregated by dowser; ● "best" choices plotted in (A)

ere is another interesting way of considering the erall performances of these six "outstanding" dows- which is shown in Fig. 4, where their errors (dis- ces between observed choice and pipe location) are mpared with what might have been achieved, if, on ry test, the dowsers had simply indicated that they ught the water pipe was located exactly in the mid- of the 10-m test line. As that presentation indi- es, choosing the midpoint would have been a rela- ly successful strategy. For five of the six dowsers, average distance between the water pipe and the middle of the test line was less than the average error actually made when dowsing, by amounts ranging from 43 to 110 cm. The only exception to this trend is Dowser # 89; his choices averaged slightly closer to the pipe (by 4 mm) than if he had consistently chosen the middle of the test line. The potential advantages that would have been provided by this simple alternative strategy suggest that concentrated searching by the dowsers with their divining rods was a waste of both time and effort.

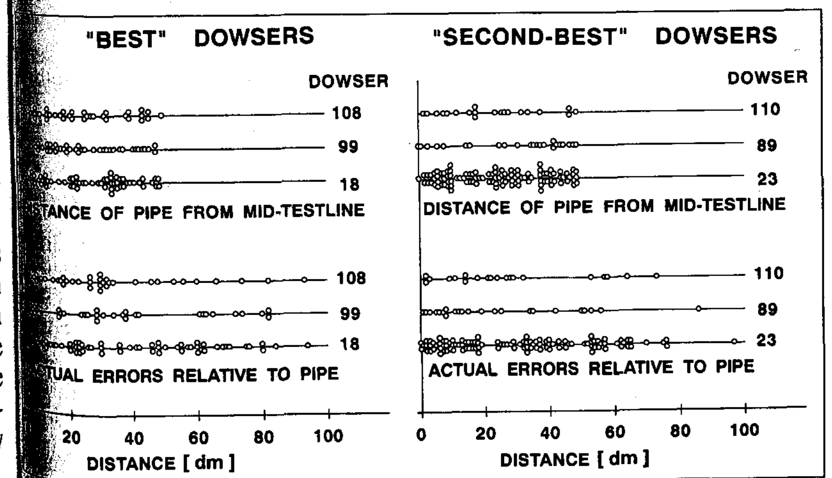

Fig. 4. *Lower three rows* of data points show magnitude of errors, relative to the pipe, made by the three "best" dowsers (*left panel*) and the three "second-best" dowsers (*right panel*). *Upper three rows* of data show how far these pipe locations were from the middle of test line in those same tests;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

Discussion

The experiments described here represent the most extensive and carefully conducted study ever undertaken to investigate the capability of dowsers to detect water at nearby, established locations. If water dowsers – even some small fraction of them – possess the ability that is claimed by so many, this study 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 below the ground, so the task here seems to involve a simpler assignment than the kinds of field problems that water dowsers regularly confront. Those who were actually tested in the final, critical experiments were preselected 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 – again 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 entire 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; "double-blind" 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. All these measures represent 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, had it demonstrated strong and reproducible skills in most of the dowsers, the outcome should be expected to serve as a springboard for intensive follow-up research by physicists as well as physologists, to explore what mechanisms might be responsible for these capabilities. Had that kind of data been obtained, then someone who remained a skeptic would be almost forced to invoke undetected cheating by the dowsers to support his position. That sort of success was not, however, achieved; as the Final Report recognizes, the vast majority of the participants in the critical experiments (who had 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 tests 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-dowsing phenomenon – that some few individuals showed an extraordinarily high success rate. The reexamination of those data described here indicates that this conclusion rests on very flimsy grounds indeed. In the Final Report, attention was focused on a small number of unusually good performances (Figs. 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 dowsing 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. 2 and 3, even the most "successful" test series were obtained by people who could not themselves replicate that sort of performance. Overall, those same "experts" did very poorly, with general success rates no better than those of the rest of the dowsers, no better than one might expect due to chance alone. Perhaps the most interesting conclusion that can be drawn from the entire analysis here is that even the dowsers, who, on single occasions, managed to perform at considerably better-than-chance levels, could on average have done better overall than they in fact did, if they had simply chosen the midpoint of the test line in each and every test.

It is my impression that the Scheunen experiments have devastating implications for the art and profes-

n of water dowsing. Although the research was conted by a group fully sympathetic to the cause of wsing, and has been interpreted by them as indicat that successful dowsing is indeed a real phenome, even the most favorable interpretation of the eximents is discouraging. The data indicate that if rets appreciably better than chance levels are to be obned, one must engage a very select dowser (one with lls better than 99% of his competitors) on one of very good days (and he cannot tell you whether it good day or not); and even then, that super-expert, his best day, is apt to be badly wrong about half the e at locating a water source that is only 3 or 4 m ay. Would you yourself be willing to pay someone his advice, if he attempted to demonstrate his comence by showing a graph resembling one of those ilrated in parts C, D, and E of Fig. 2 and 3? The rets shown in Fig. 4 could be interpreted as suggesting t in the absence of other information, it would be etter strategy, for someone who is planning to drill ell, to sink his hole right in the middle of the avail e region, rather than to rely on advice from a dowsno matter how successful the dowser may claim to e been on certain past occasions. This interpreta depends, of course, on the assumption that there single best place to drill, and that the closer one o that location, the better the expected outcome.

nclusion

ause the Scheunen experiments involved such a e-scale test program, which incorporated both erous allowances favoring the dowsing hypothesis a careful, rigorous experimental program, a defin e answer is finally available to the central, age-old stions about water dowsing. Briefly stated, the clusion is that even with very extensive testing, by archers sympathetic to the cause, no persuasive dence could be found for reproducibility of the wsing phenomenon", neither interindividual reproibility nor intraindividual reproducibility. Instead, as now been demonstrated that:

1. *If* the ability to locate water from a distance by extraordinary stimuli exists, that skill cannot be reproducibly demonstrated across a select group of 43 experts from among 500 dowsers who all *think* hat they have the ability (a conclusion consistent ith the summary in the Final Report);
2. In those few cases in which a single series of tests suggests that a given dowser may perhaps have better-than-chance abilities, similarly good results are not reproducible by that same individual in other 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 extensive the program? Thus, 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.

This does not, of course, constitute a rigorous refutation of the dowsing 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 unusually talented individuals exist, distinguishing them from among the unskilled appears to be a hopeless task. This leads to a valuable insight: whether one prefers the interpretation that truly skilled dowsers exist, who are so rare that none was found in the Scheunen experiments, or whether one instead prefers the interpretation that the ability claimed by dowsers 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 93-13038 from the National Science Foundation. Valuable comments on the manuscript were made by Dr. E. Vetter.

1. Hyman, R., Vogt, E.Z.: Psychol. Today *1* (11), 35 (1967)
2. Ellis, A.J.: Water-Supply Paper 416, Department of the Interior, U.S. Geological Survey. Washington, D.C.: Government Printing Office 1917
3. Wagner, H., Betz, H.-D., König, H.L.: Schlußbericht 01 KB 8602. BMFT 1990
4. Erickson, B.H., Nosanchuk, T.A.: Understanding Data. Toronto, Buffalo: Univ. of Toronto Press 1992
5. Mosteller, F., Tukey, J.W.: Data Analysis and Regression. Reading, MA: Addison-Wesley 1977.
6. Tukey, J.W.: Exploratory Data Analysis. Reading, MA: Addison-Wesley 1977