

Empirical Evidence for a Non-Classical Experimenter Effect: An Experimental, Double-Blind Investigation of Unconventional Information Transfer

HARALD WALACH AND STEFAN SCHMIDT

*Department of Psychology, University of Freiburg, Rehabilitation Psychology, D-79085 Freiburg,
Germany*

Abstract — We set up a rigidly controlled, double-blind dowsing experiment with three repetitions to test whether dowsers are able to extract information out of a system in an unconventional way. One hundred and four professional and lay dowsers had to distinguish between randomly distributed, sealed and indistinguishable probes of pure mineral water or parathione, using a one-hand dowsing rod. The subjects were unable, on the whole, to distinguish between the probes better than chance. Performance was significantly negatively correlated with paranormal beliefs. Subjects instructed by one among three blinded experimenters were able to distinguish between the probes significantly better than chance. As we have excluded any conceivable way of leakage of relevant information, we conclude that we found a non-classical experimenter effect.

1. Introduction

Dowsers and users of pendulums typically claim to be able to extract information from physical systems without any known causal pathway. This can be geological information (Betz, 1995), information on missed or distant objects or persons, information on suitability of food, or the like. It was this claim which we tested, since it is relevant to know whether there are pathways of information transfer other than known causal pathways. While unusual water prospection and other geological dowsing phenomena could possibly be explained by specific electromagnetic and corresponding neural sensitivity, other kinds of information extraction could, if proven, rest on unknown pathways. Earlier experiments have yielded inconclusive results. While the issue of dowsing for water has been investigated recently with promising results (Betz, 1995), the question of unconventional information transfer on the whole is far from settled.

Generally, tightly controlled experiments were unable to find evidence for unconventional information transfer (Hansen, 1982) in dowsing and pendulum experiments. In the tightly controlled experiment reported here, the question was whether professional and lay dowsers are able to extract information out of a system under blind and controlled conditions, using a commercially

available one-hand rod, which is sold for its easy-to-use properties and well known in German speaking countries.

Throughout this article we use two terms in different contexts: We speak of "unconventional" in the context of information transfer, thereby suggesting that information retrieval does not occur by known physical or sensory channels. We use the notion "non-classical" when we speak about the influence of an experimenter on an experimental system, distinguishing this effect from the known Rosenthal effect (Rosenthal, 1976), which is mediated by sensory, albeit subliminal, channels.

2. Methods

Material and Procedures

The commercially available one-hand rod, developed by Erich Korbler, that we used is claimed to be generally usable and applicable to all sorts of dowsing problems (*Raum & Zeit*, 1995). Its use allegedly requires very little training and no specific dowsing faculties. It is an elastic plastic rod, 58 cm long, with a wooden ball (2 cm diam.) attached to its top end, and a plastic handle. All relevant steps of the experiment and the material used were checked with dowsing experts, named by the producer of the rod. Ten probes of the purest known mineral water (Volvic, 2.5 ml) and ten probes with parathion, a once widely used, now prohibited insecticide, (E 605, 0.2 g dissolved in distilled water to 2.5 ml, a lethal dose) were placed in quartz glass vials (45 mm, 11 mm diam.) with silicone stoppers, wrapped in wild silk, and put in brass containers (100 mm, 16 mm diam.) soldered to brass stands (100mm × 20mm × 1mm), which were sealed with a rubber stopper, and a paper seal. The paper seal was strongly adhesive and gave the date of preparation. Tests showed that the seal could not be detached without leaving traces.

The probes were randomized using a real random number code stemming from a Zener-diode. They were fixed between two wooden boards such that they could only be touched and not handled in any other way. The boards were screwed together from the downside, and firmly taped on tables using double-sided tape which is usually used for fixing carpets. Every experimental testing was videotaped to make sure that subjects adhered to the protocol, which allowed them only to touch the probes, not to handle or open them. The subjects neither knew the contents of the probes nor the number of respective probes. The only information they had was that they had to distinguish 20 probes as to whether they would have a positive or negative effect on their organism. We had the experimental room, a quiet seminar room on the first floor in the building of the Department of Psychology at the University of Freiburg, checked by a professional dowser for "ley lines", to make sure that purported geological signals would not interfere with the experiment. The probes were all placed outside so-called geopathic zones. The experimental room was locked after each experimental session and only opened for experimental trials. The trials

were carried out during summer vacation, when the building generally is quiet and not highly populated.

Every subject filled in a questionnaire, asking for basic sociodemographic data, and information on experience with dowsing rods and pendulums, theoretical and practical knowledge, and professional use of dowsing. Part of the questionnaire was a scale of paranormal beliefs, which had been constructed, tested, and validated in a different study (Brednich, 1993) with four to six items for each of the following subscales for belief in "magical forces", "astrology", "ESP and precognition", and "reincarnation".

Design

We carried out three independent experiments with approximately 35 subjects per experiment. Every experiment was randomized independently, the code was locked; the randomization procedure was videotaped. Each experiment took three successive days in a week. Experiment 1 and 2 were on two successive weeks in August 1995, experiment 3 followed one week later. Since we also wanted to test the specific hypothesis that the information gained in an experimental setting reduces the magnitude of the effect size discovered in a repetition of this experiment (Lucadou, 1995; Model of Pragmatic Information), it had been specified in the protocol that the data of experiment 1 and 2 should be evaluated together after experiment 2, and the pooled data at the end. Therefore, we evaluated experiments 1 and 2 after experiment 2 was finished, and randomized experiment 3 with the information obtained by the two preceding experiments. Before the beginning of the experiments, the protocol was deposited with Eberhard Bauer, the editor of the *Zeitschrift für Parapsychologie und Grenzgebiete der Psychologie*, the German Parapsychology Journal.

Subjects and Experimenters

One hundred and twenty five volunteers, 20 professional dowsers as well as interested lay persons, responded to a publicity campaign in the press and local media, 111 attended four experiments, 106 gave data. Two persons had missing data of more than a quarter and therefore were excluded, leaving the final number at 104. Volunteers came from all over Germany at their own expense and participated without reward or reimbursement. Fifty seven (54%) were female, 47 were male, with a mean age of 44.1 years ($sd = 13.3$; range 22-76 years).

Different individuals with different degrees of information about the design and sequence of probes were involved with the experiment. The supervisor (HW) knew about the design, set up the randomized sequences, kept the code in a locker, and did not interfere with the actual experimental procedure during the whole experiment. He was in a separate building about 500 yards away from the experimental rooms, unknown to the subjects. The senior experi-

menter (SS) knew about the design, but was blind to the actual sequence. He resided in a room in the third floor of the building, and saw the experimental subjects only after they had deposited their data, when they were invited to come up to the third floor to relate their experience. Three hired junior experimenters, one graduate psychology student, a teacher, and a social worker, knew nothing about the actual design, except that probes had to be distinguished as positive or negative using a rod, and were also blind to the randomized sequence. They had to explain the task to the subjects in a training session in a different room, on the ground floor of the building, next to the entrance, before the commencement of the experiment, using open test material and written information.

Procedure

Subjects were scheduled by phone to arrive one by one. When subjects arrived for the experiment, they were greeted by one of the junior experimenters, and brought into the waiting room. Here they could read introductory material, learn how to handle the rod, and use it on different specimens. Having learned about the general purpose of the experiment, namely, to find out whether something unknown to them was "good" or "bad" for them, subjects were taught how to interpret the different movements of the rod: They were instructed to close the eyes and to recreate the memory of a very pleasant situation, to remember the feeling, feel an inner "yes", and then to observe the movement of the rod. After that they were instructed to remember an unpleasant situation, feel an inner "no", and again observe the movement of the rod. This was repeated until each subject was sure and felt that he or she knew exactly the idiosyncratic "yes" and "no" sign of the rod. They could then try this experience on open material, like a rotten and a good tomato, water with a lot of, and a little, salt, and to specimens of the sealed probes to be tested afterwards, marked "+" and "-". Thus subjects learned to "gauge" the movements of the rods to receive clear "yes" and "no" "answers" from the rod. Experimenters were instructed to leave as much time to the subjects as they wished. Subjects only proceeded to the testing, when they felt sure that they could handle the rod. The training was deemed sufficient, since nearly all subjects described the task in the experimental room as easy and clear. After training, subjects proceeded to fill in the questionnaire. Specific care was taken that experimental subjects would not meet each other between trials. They were then guided to the experimental room, the video camera was switched on, and the subjects were left alone for as long as they needed to do their testing.

The task was to find out whether the probe was "good for me" or not, and to tick a box on the data sheet with a plus or minus sign accordingly. The data sheets were then sealed and put in a ballot box and were not handled until the experiment was finished. The ballot box was a box normally used for University elections, locked with a key which was held by the senior experimenter (SS). It was left in the locked room with the data in it, until the end of the sec-

ond experimental week, and again for the last week. After subjects had finished their testing and deposited their data in the ballot box, they were invited to come upstairs for a drink to the third floor in order to give them the opportunity to meet the experimenter (SS) and to keep them out of reach of other persons at the same time. Participants generally had a very positive attitude towards the experiment and found the task easy and clear-cut.

Evaluation

Coded data were double checked against the data sheets. Number of hits ("+" with water and "-" with parathion) and number of misses were compared using Wilcoxon Signed Rank tests for paired data. When testing for hypotheses stated in the protocol, we used one-sided p-values for apriori hypotheses. When, on the other hand, we tested for exploratory hypotheses, we used two-sided p-values for exploratory analysis. Correlations of hit rates and other variables recorded before the experiments — professionalism of dowsing, paranormal beliefs — were evaluated by Spearman's correlation. All calculations were carried out using SPSS for Windows.

3. Results

The expected chance rate of 10 hits out of 20 was not exceeded under strictly double-blind conditions ($\bar{x} = 10.16$, $sd = 1.98$, range 6 – 15). The protocol called for an evaluation of the data after experiment 2 (see section 2). When, at that point, we further explored and broke down the data according to different experimenters, however, subjects instructed by experimenter 3 had a hit rate significantly above chance ($p = 0.04$, Wilcoxon, two-sided) for experiments 1 and 2 together. We subsequently checked, whether this significance would be repeatable by testing an additional hypothesis introduced after experiment 2 stating that there will be no differential effects of experimenters. Experiment 3 in itself was not significant due to a lack of statistical power, but showed the same tendency. Nevertheless, the overall result was stable: Subjects introduced to the experiment by experimenter 3 had a hit rate significantly higher than chance ($p \leq 0.01$, Wilcoxon, one-sided). Details are given in Table 1.

There was a significant negative correlation between paranormal belief and hit rate ($r = -.29$; $p = .002$, two-sided). Therefore we blocked this variable into

TABLE 1

	Experiment 1 & 2	Experiment 3	All Experiments
Experimenter 1	9.86 (1.83) [22]	9.00 (2.00) [9]	9.61 (1.89) [31]
Experimenter 2	9.91 (1.95) [22]	10.13 (2.00) [15]	10.00 (1.94) [37]
Experimenter 3	10.96 (2.13) [25]*	10.45 (1.57) [11]	10.81 (1.97) [36]**
Total	10.28 (2.02) [69]	9.94 (1.91) [35]	10.16 (1.98) [104]

Mean hit rate (standard deviation) [n] globally and per experimenter, as evaluated (*i.e.* experiments 1 & 2, experiment 3, all), *: $p \leq 0.04$ (Wilcoxon, two-sided), **: $p \leq 0.01$ (Wilcoxon, one-sided)

the categories high, medium and low and checked visually, whether by chance experimenter 3 had most of the subjects low in paranormal belief. This was not the case. Subjects with experimenter 3 had higher hit rates independent of belief (Figure 1).

Profession, age and sex were not significantly correlated with hit rate. The effect size for the resulting experimenter effect is $d = 0.5$, a medium size effect in Cohen's terminology: $[x_3 - (x_1 + x_2)/2]/sd$ (Cohen, 1987). We had an approximate power of $\beta = 0.60$ to discover the effect.

The negative correlation between hit-rate and paranormal belief can be considered as a potential "signal" in the sense of Lucadou (1995). For the first experiment, this correlation was $-.26$ ($p = .14$), for the second $-.44$ ($p = .007$), and for the third $-.19$ ($p = .28$). This means that the correlation increases with experiment 2, and decreases at experiment 3. A meta-analysis according to Rosenthal (1991), *i.e.* transforming the correlation coefficients and testing for homogeneity, yields the result that all three correlations, as well as the correlation of experiment 3 against the mean correlation of experiment 1 and 2 together, belong to the same population of correlations. We have, therefore, not found conclusive statistical evidence for a decline-effect between experiments, although this might be suggested by visual inspection of a corresponding plot.

4. Discussion

The statistical analysis of the data shows no evidence that subjects under strictly double blind conditions are generally able to extract information out of a system using a dowsing rod, neither as a group nor in individual performance. It is hard to dispute this result, since we tried to enhance external validity as much as possible. The experiment was praised by experts, named by the

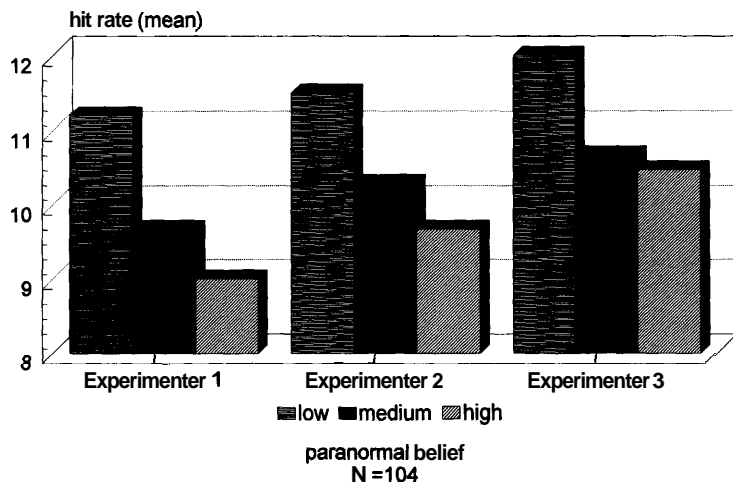


Fig. 1. Hit rate by experimenters and by paranormal beliefs low, medium, and high.

producer of the rod, as valid, before we started. Volvic water and parathion are universally positive or negative for people. No known allergy or medical condition exists which speaks against Volvic water, and, given the normal rates for suicidal inclinations, parathion could have been subjectively "good" for at most 2 persons in our sample. For the others it would have been lethal and therefore bad. Besides, in using parathion we employed a substance which in sub-toxic doses is widely distributed in the ecological chain due to its former use as insecticide. Shielding and interference of probes was checked with experts. Feedback of the participants regarding the experimental in general was favorable. A point for debate would be that all subjects had to use the same rod. This could have compromised the "dowsing faculty" of experienced dowsers. However, the rod used in our experiment is said to be universally applicable and easy to use by everyone, and is explicitly sold as such. We did not receive negative feedback about the rod except indirectly by the seven people who did not finish the experiment or did not want to use the rod. This dropout-rate of 6% can hardly be a hint that the rod might have been a problem.

As we kept a strictly blind regimen and covered the whole experimental process with videotaping, there was no overt or covert information whatsoever which could have been conveyed by the experimenters to the subjects. Experimenters did not know contents or ratios of probes, nor the actual sequence, nor hypotheses. All they knew was that a rod is being tested using probes that would be "good" or "bad" for people. Their only interaction with experimental subjects took place in the instruction phase of the experiment, by showing and explaining the rod, and helping the subject to get acquainted with the task. After they had switched on the video tape and closed the door of the experimental room, they had no contact with subjects and their data any more. By keeping the data in a ballot box within the experimental room, which was locked all the time, we made sure that the data would not be handled during the experiment.

Against this background of rigid control, it is surprising that subjects were able to distinguish significantly better than chance only when instructed by one specific experimenter. This experimenter effect is small but significant and consistent within the whole set of experiments: We found this result when evaluating experiments 1 and 2 with a two-sided test by chance. We then formulated a hypothesis, and did in fact not expect to see it again after the third experiment. It became stronger after all three experiments, due to more statistical power. We did four tests; one for the group as a whole and three for each experimenter, using very robust statistics. The result remains significant even when corrected for multiple testing. Given the fact that we had roughly 35 subjects per experimenter, ten per week and experimenter, and that the effect remained stable, this is a puzzling result, which cannot be considered as an experimental artifact. The negative correlation between paranormal belief and hit rate does not explain it. There was no weekday or serial effect, as every experimenter participated in every experiment, but in a different sequence. Nei-

ther did one of the three experiments in itself show a significant effect of such a kind. We also checked for position effects and did not find any.

It could be the case that with one experimenter the subjects felt more at ease, thus being able to employ a capacity to extract information unconventionally, whereas with the other experimenters subjects felt more tense. We do not have data on the way subjects perceived experimenters. But even if there were a remarkable difference in this sense, this would mean that a capacity to function in an unconventional way can be triggered or catalyzed by a person not himself familiar with the task itself.

So called psi-experimenter effects are well-known in the literature (Kennedy & Taddonio, 1976). They refer to experimenter influence which is not mediated by known sensory channels, in contrast to classical experimenter effects, which are supposed to work by subliminal perception and hidden cues (Rosenthal, 1976). We saw in a comparatively simple, tightly controlled experiment that an experimenter clearly modifies the result of a task by unconventional means. He does this without any information about the target and the task to be performed. We conducted informal interviews with the experimenters after the study, asking them about their attitudes and expectations. All of them were positive about the experiment without expecting particular results. The experimenter in question (the student) described himself as interested, having come to believe in dowsing, and hoped the experiment would prove the dowsing phenomenon, the chance of which he rated at 50:50. The other experimenters had also positive attitudes towards the experiment, one being confident to see an effect, the other not. There is no striking difference between the experimenters, except that the attitude of experimenter 3 was a little more enthusiastic, without, however, being strongly biased towards a specific outcome. None of the experimenters had any relevant information whatsoever to be conveyed.

What has been found for our quite artificial system "probe-rod-dowser-experimenter" could be the case for any system under experimental control, and even mores so for uncontrolled systems. The experimental doctrine according to which it is only the actual information about a system that matters could well be too shortsighted. Our results suggest that not only the information that an experimenter has about an experimental system is relevant. Perhaps also other psychological variables, which we do not know yet, can, in hitherto unknown ways, modify experimental outcomes. Due to often heard, yet rarely officially published scientific gossip some labs are more lucky than others in reproducing or finding results in different areas of science. This quasi-fact might have its root in the observation that a specific experimenter can modify the outcome.

This result needs repetition. Our third experiment can be seen as a first positive experimental *a priori* test of the hypothesis of a non-classical experimenter effect. Moreover, we do not know which aspects of the experimenter's personality are vital to this purported effect. Apart from that, we suggest that

investigators not only report blinding procedures in experiments, but also describe circumstances which might have a bearing on the results: importance of the experiment to one's career, expectation, motivation, attitudes, and the like. Our results show that the cut which is usually assumed between the "mental world" of our concepts and the "material world" out there, between models and facts (Atmanspacher, 1994) is not as thorough and stringent, as some might wish. They also show that under certain circumstances, which need clarification, unconventional "transfer" of information can occur.

Another effect exhibited by the data was the strong negative correlation between paranormal belief and hit rate: the higher the belief, the poorer the outcome. Contrary to expectation, professionalism does not correlate with hit rate. This could be an indication that "believers", who might have put pressure on themselves to "prove" a paranormal fact such as unconventional information transfer, performed poorly because of their self-generated tension. This suggests that a "playful" attitude is favorable for demonstrating paranormal effects.

Lucadou's hypothesis that the effect size for paranormal phenomena decreases between successive experiments if there is a consistent effect which could be used to code a signal, as soon as the effect is known, was not vindicated by our experiment. Assuming that the negative correlation between paranormal belief and hit rate represents a potential "signal" in the sense of Lucadou, our experimental results remained inconclusive with respect to a corresponding decline effect. Presumably one would have to use larger databases to actually model the effect. As it is, we have a qualitative hint that the hypothesis might be valid, but we have no clear quantitative evidence from our experiment.

Acknowledgements

We gratefully acknowledge financial support by Ehlers-Verlag, Hans-Joachim Ehlers, for the setup of the experiment, logistic support by Dr. Joop M. Houtkooper, who provided the random numbers, Dipl. Psych. Anke Brednich, who helped with installing the probes, and Dipl. Psych. Eberhard Bauer, Dr. Dr. Walter von Lucadou, Prof. H.D. Betz, Prof. J. Mischo, Dipl. Ing. Karsten Krause, and Ruth Timter, who gave helpful advice in planning the experiment. We thank our three experimenters Christina, Thomas and Gerd, and all the participants. We gratefully acknowledge helpful criticisms by Dr. Harald Atmanspacher.

References

- Atmanspacher, H. (1994). Complexity and meaning as a bridge across the Cartesian cut. *Journal of Consciousness Studies* 1, 168.
- Betz, H.-D. (1995). Unconventional water detection: Field test of the dowsing technique in dry zones: part 1 & 2. *Journal of Scientific Exploration*, 9, 1-43; 159-189.

- Brednich, A. (1993). *Eine Fragebogenuntersuchung zur Erfassung von magisch-irrationalen Denken und der Schizotypischen Persönlichkeitsstörung bei Erwachsenen*. University of Freiburg, Department of Psychology, unpublished thesis.
- Cohen, J. (1987). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale: Lawrence Erlbaum.
- Hansen, G. P. (1982). Dowsing: a review of experimental research. *Journal of the Society for Psychological Research*, 51, 345.
- Kennedy, J. E. & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology*, 40, 1.
- Lucadou, W. v. (1995). The model of pragmatic information (MPI). *European Journal of Parapsychology*, 11, 58.
- Raum & Zeit, Special 3 (1995). Die Sprache der Natur verstehen. *Das Lebenswerk Erich Körblers*. Sauerlach: Ehlers Verlag.
- Rosenthal, R. (1976). *Experimenter Effects in Behavioral Research (enlarged edition)*. New York: Irvington.
- Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*, Newbury Park: Sage.