A PRELIMINARY STUDY OF THE EFFECT OF DATA DESTRUCTION ON THE INFLUENCE OF FUTURE OBSERVERS

By Dick J. Bierman and Debra H. Weiner

ABSTRACT: Observational theories predict that the results of psi tests can be influenced by multiple observers of the data including future observers. This experiment tested a modification of these theories here called the potential-observer theory, which states that the degree of a future observer's influence is related to the probability that he or she will actually observe the results.

The subject (D. J. B.) completed a total of 120 512-trial PK runs (P = 1/2) with trial-by-trial feedback. The data of each run were divided by a fixed method into four subruns which were then subjected to a process whereby a random decision determined for each data point whether or not it would be destroyed. By varying the probability of destruction across subruns the probability of future observation and thus the potential influence of future observers, was manipulated.

A future-observer effect was sought by arbitrarily dividing the data into two sets and having each author analyze (i.e. observe) one of the sets. Differences between the results of the two sets could be interpreted as an effect of future observation. The authors predicted that analyzer (A) differences would vary with destruction rate (DR) leading to a significant interaction between the two variables. It was also predicted that the results would increase in significance with increasing DR.

Four-way analyses of variance (A x DR x session half x days) were carried out separately on direct PK scores and random PK (deviations from MCE regardless of target direction). Though the results for direct PK were not significant the A x DR interaction for random PK was significant. However the obtained interaction was not in the predicted form since significant analyzer differences were found at the higher destruction rates so the theory is contraindicated. The second hypothesis was not supported. The results are discussed in terms of evidence for future-observer effects.

The "observational theories" of Walker (1975) and Schmidt (1975) predict that a random event can be affected by individuals observing the outcome, including persons who will observe it in the future. How the psi influences from various sources combine to affect the event, and related questions, have received some theoretical attention (Hartwell, 1977; Houtkooper, 1977; Millar and Hartwell, 1979; Schmidt, 1975, 1977, 1978; Walker, 1977).

As a solution to the problem of the effect of future observers, Millar and Hartwell (1979) proposed that in any experiment all humankind may be participants to a degree that depends on the *probability* of the individual's later becoming an actual observer of the outcome of the experiment.

The theory is not explicit as to how to determine this probability. It is essential, though, to define a point in the history of the experiment at which the value for this probability is determined. This might be the moment when the design is evolving in the mind of the researcher, which seems to be the position that Millar and Hartwell take in their paper. However, such an operationalization yields a cyclical logic, since the experimenter's creative processes might contain random elements and might therefore be susceptible to psi, too. Alternative points might be the moment that the data are "registered" on a "macroscopic" medium or the moment that the data are first observed consciously (generally by the subject). A second weak point in the paper is that no quantitative relationship is given between this "probability of becoming an observer" and the actual magnitude of the contribution of this potential observer to the final result.

Even though this modified observer theory, or the potential observer theory as we will call it, is not developed in sufficient detail, one of its implications is quite specific and testable. Consider as an example for such a test the following three-step experiment:

1. A large number, N, of binary random events is generated and recorded while the subject makes a PK effort to obtain more "heads" than "tails." The subject receives full feedback of the outcome of each event.
2. From these *N* data a certain fraction, *q, is* randomly selected and destroyed. This is done automatically such that the destroyed data are observed only once, by the subject, during the test run.
3. The remaining (1 - *q) N* events are inspected and analyzed later by the experimenter/analyzer.

Thus, for each random event in the test run, the probability of its being observed by the subject is 100% while the probability is reduced with regard to observation by a future observer, the experimenter/ analyzer. For this test situation, the potential-observer theory makes the surprising prediction that, provided experimenter/analyzer observation plays a vital role in the outcome, the observed scoring rate should be dependent on the destruction rate *q*.

In a recent fast RNG PK study (Bierman & Houtkooper, in press), part of the data was destroyed as described above, with a destruction rate, q, of 50% in the pilot series and of 75% in the confirmation series. Under the assumption of the potential-observer theory, this procedure should reduce the probability of future observation and hence should localize the psi source more closely to the subjects. Only marginal psi occurred in the study, which made a clear evaluation of the potential-observer model virtually impossible. However, a reduction of analyzer effects from pilot to confirmation series justified further exploration of the model.

METHOD

The subject (D. J. B.) did 20 runs of fast RNG PK per day on six consecutive days in July of 1979. Each run consisted of 512 trials (P = 1/2). Performance of the RNG was displayed graphically on a trial-by trial basis, so the subject observed all generated trials with a probability of 100%. After each session the data of each run were subject to destruction following a fixed and predetermined algorithm. The 1st, 5th, 9th, etc., trials were left undisturbed, resulting in a subrun of 128 trials with an MCE of 64 hits. An RNG decision whether or not to destroy the result of each of the 2nd, 6th, 10th . . . trials was made randomly, the probability of destruction being equal to .50, which resulted in a subrun of an average of 64 trials. Trials numbered 3, 7, 11 . . . were subjected to a similar destruction process, except that the probability of destruction was .75. The remaining trials were destroyed with a probability of .875. Thus, each complete run finally resulted in four subruns with destruction rates of 0%, 50%, 75%, and 87.5%, respectively. After completion of the experiment, the data pool consisted of 80 subruns per day of testing. These data were transferred from the DECLAB computer in Amsterdam to the Institute for Parapsychology's PDP 11/45 for further analysis.

It was decided that the data from the odd days were to be analyzed by D. J. B., while the data of the even days were to be analyzed by D. H. W.. The reason for this decision was that a split-data analysis with these two analyzers using previously observed data had yielded analyzer differences which could be interpreted as future-observer effects (Weiner & Bierman, 1979). The question in the present experiment was whether or not the partial destruction of the data before analysis would result in a decrease of the analyzer effects.

Raw scores were normalized into z scores (without continuity corrections) in order to compensate for varying numbers of trials among the four destruction-rate subruns[1]. Analyses were carried out separately for the four destruction rates and separately by day of self-test. Each analyzer had free rein regarding the order

---

[1] The transformation treated the destruction probability as part of the Ph thus, the transformation of the 50% destruction rate data calculated the z score on te basis of 128 trials with P = .25 (Pdes x Phjt) and so on for the remaining destruction rates.

in which the tests were to be done. After the analyzers had finished working with (i.e., observing) their portion of the data, they met and compared results. The formal statistical tests described below were then carried

HYPOTHESES

1. An analyzer-by-destruction-rate (A x DR) interaction was predicted for the variables:
    a. Direct PK
    b. Random PK (variance from MCE).
2. It was predicted that results pooled over analyzers would increase in significance with increasing destruction rate.

ANALYSES

The planned hypotheses were tested by four-way analyses of variance conducted separately on the two dependent variables. Besides the analyzer and destruction-rate factors, it was decided that the analysis should include the factor "days" (nested within the analyzer variable) as a control measure in order to verify that any effect of destruction rate was not due to the results of a particular day or days of testing. It was also decided (after some observation of the data) to include session half as a post hoc factor. The analyzer and days/analyzer factors were considered random independent variables while destruction rate and session half were fixed[1]. Hypothesis 1 predicts a significant A x DR interaction, while Hypothesis 2 would be supported by a significant destruction rate main effect.

Data Transformations

As mentioned above, raw scores were normalized in order to make comparable the data of the four destruction-rate subruns. Random PK was calculated as the square of the direct PK *z* score. This measure is distributed as chi-square with one degree of freedom and thus is highly skewed. In order to render it approximately normally distributed, a method derived by Wilson and Hilferty (1931) was applied.

RESULTS

Direct PK

The results shown in Table 1 reveal no support for either hypothesis[2].

---

[1] This model entails changes from the more commonly used fixed model in the selection of error terms to evaluate certain factors (Keppel 1973). Of relevance to this paper is the fact that the A x DR interaction is tested against the DR x days/analyzer mean square and that the A x DR x session half interaction is tested against the DR x sesslon half x days/analyzer mean square

[2] The Levene test (Keppel 1973) was carried out on these data and showed significant heterogeneity of variance in the A x DR interaction and across days. Though this result violates the assumption of the ANOVA and might render it inappropriate for a test of direct PK it has been shown that for equal sample sizes variance difterences on the order of l:much greater than the ratios in the present data—do not seriously affect the validity of the t and *F* tests (Boneau 1960/1971). On a practical level the effect of heterogeneity is of particular concern when the obtained *F* ratio is close to the critical value for

Table 1
ANALYSIS OF VARIANCE FOR DIRECT PK

Source	Sum of Squares	*df*	Mean Square	F
Analyzer[a]	.04	1	.04	< 1
Destruction rate	1.38	3	.46	1.39
Session half[b]	.07	1	.07	< 1
Days/analyzer[a]	.29	4	.07	< 1
A X DR	.99	3	.33	< 1
A X session halfb	.21	1	.21	< 1
DR x session halfb	1.06	3	.35	1.21
DR x days/analyzer	6.42	12	.54	< 1
Session half x days/analyzer	5.56	4	1.39	1.30
A x DR x session halfb	.88	3	.29	<1
DR x session half x days/analyzerb	9.74	12	.81	< 1
Within groups	462.22	432	1.07	
Total	488.87	479		

[a] Analyzer and days/analyzer are random independent variables; see Footnote 2.

[b] Session-half factor is post hoc.

---

significance. The *F* ratios in the direct PK ANOVA are well below the significance criteria which makes the problem of heterogeneity of less consequence. The analysis of random PK or variance from the theoretical mean is related to the Levene test which investigates variance about the empirical mean; hence the results of the Levene test roughly paralleled those described in the next section of the paper

Table 2
ANALYSIS OF VARIANCE FOR RANDOM PK

Source	Sum of Squares	*df*	Mean Square	F
Analyzer[a]	.025	1	.025	<1
Destruction rate	1.499	3	.50	<1
Session half[b]	1.261	1	1.261	<1
Days/analyzer[a]	11.887	4	2.972	3.20**
A x DR	12.773	3	4.258	6.71***
A x session half[b]	1.687	1	1.687	< 1
DR x session half[b]	.421	3	.140	< 1
DR x days/analyzer	7.624	12	.635	< I
Session half x days/analyzer[b]	8.867	4	2.217	2.39*
A x DR x session half[b]	7.224	3	2.408	5.52**
DR x session half x days/analyzer[b]	5.229	12	.436	< 1
Within groups	401.372	432	.929	
Total	459.869	479		

[a] Analyzer and days/analyzer are random independent variables; see Footnote 2
[b] Session-half factor is post hoc.
*p < .05.
**p < 025.
***p < .01.

Random PK

The results of the analysis of variance on random PK scores are shown in Table 2. Although Hypothesis 2 was not supported, the predicted A x DR interaction is significant at the p< .01 level, F(3,12) = 6.71. Analyses of the simple main effects reveal that the interaction is caused by suggestive differences between analyzers at the 0% and the 75% destruction rates, and a significant difference at the 87.5o destruction rate, F (1,432) = 6.98, p <.01. Further, the DR variable shows significant differences in D.J.B.'s data, F(3,432) = 3.20,p<.025, but not in D. H. W.'s data.[1] (See Table 3.)

---

[1] If a future observer has an influence on the data, then psychological factors (e.g. mood and expectation) in this individual are as important to the outcome as similar factors in the subject. Before observing her portion of the data D. H. W. wrote out her expectations regarding the outcome. Analyses to assess these expectations (not reported here) indicated some agreement with the results; for example it was expected that DR would be more likely to show an effect in D. J. B. s data than in D. H W. s.

Table 3
SIMPLE MAIN EFFECTS OF A*DR INTERACTION IN RANDOM PK

Analyzer Over Levels of Destruction Rate

Source	Sum of Squares	*df*	Mean Square	F
Analyzer at 0%	3.089	1	3.089	3.33
Analyzer at 50%	.351	1	.351	< 1
Analyzer at 75%	2.876	1	2.876	3.10
Analyzer at 87.5%	6.481	1	6.481	6.98**
Within groups	401.372	432	.929	

Destruction Rate Over Levels of Analyzer

DR for D.J.B.	8.915	3	2.972	3.20*
DR for D.H.W.	5.357	3	1.786	1.92
Within groups	401.372	432	.929	

*p < .025.
**p < .01.

The analysis of variance also yielded a difference in random PK performance among the days of testing, F (4,432) = 3.20, p < .025. In order to isolate the components of this result, a Scheme test was applied, which showed a significant difference in the subject's random PK performance between two particular days (p < .05). This was caused by scores too close to MCE on one day and scores too distant from MCE (i.e., exhibiting both psi-hitting and psi-missing) on the other. It is important to note that these results do not intrude upon the A x DR interaction. Both of the days of extreme performance formed the data set for D. J. B., so the effects are cancelled when combined for the analyses of interest; More importantly, it is the DR x days/analyzer interaction that would confound the significant A x DR result; not only is the DR x days/analyzer interaction not significant, any such interaction is taken into account when determining the significance of A x DR by treating days/analyzer as a random independent variable. (See Footnote 2.)

Table 4
SIMPLE INTERACTION EFFECTS OF THE ANALYZER X DESTRUCTION RATE
X SESSION HALF INTERACTION IN RANDOM PK (POST HOC)

Source	Sum of Squares	*df*	Mean Square	F
A x DR: First half	11.669	3	3.890	4.19**
A x DR: Second half	8.328	3	2.776	2.99*
Within groups	401.372	432	.929	

*p < .05
**p < .01

Session Half (Post Hoc)

During the observation analyses D. J. B. noted differences in random PK performance between the first and second halves of the sessions. This variable was included in the ANOVA as a post hoc factor. As indicated in Table 2, the A x DR interaction varies significantly with session half. Further investigation revealed that while the interaction is significant in both halves considered independently, it is stronger in the first half (Table 4). In the first half, significant analyzer differences are found at the 75% and 87.5% destruction rates while, in the second half, a significant difference is found at the 0% destruction rate (Table 5). The DR variable is significant for D. J. B.'s data in the first half, $F\ (3,432) = 2.87$, $p < .05$, and is only suggestive in the second half.

Session half also interacted with the differences in random PK performance among the days of testing. The difference among the days is concentrated in the first half of the sessions, $F\ (4,432) = 4.39$, $p < .01$, and does not appear in the second half.

Table 5
SIMPLE, SIMPLE MAIN EFFECTS OF ANALYZER X DESTRUCTION RATE X
SESSION HALF INTERACTION IN RANDOM PK (POST HOC)

Source	Sum of Squares	*df*	Mean Square	F
A. First Half:				
Analyzer at 0%	.272	1	.272	<1
Analyzer at 50%	.466	1	.466	<1
Analyzer at 75%	8.060	1	8.060	8.68**
Analyzer at 87.5%	3.933	1	3.933	4.23*
B. Second Half:				
Analyzer at 0%	3.858	1	3.858	4.15*
Analyzer at 50%	2.312	1	2.312	2.49
Analyzer at 75 %	.194	1	.194	<
Analyzer at 87.5%	2.616	1	2.616	2.82
Within groups	401.372	432	.929	

*p < .05.
**p < .01.

DISCUSSION

The predicted analyzer x destruction-rate interaction was significant for random PK; however, the interaction was not in the form expected from the potential-observer model. The anticipated interaction would have been caused by a decrease in analyzer differences at the higher destruction rates, but in the present study significant analyzer differences were found at these destruction rates. Thus, the results contra indicate the model.

It may be possible that the present destruction process does not adequately screen the data from the effects of future observers and that a second level of destruction (between the "observation by analysis" stage and the planned ANOVA'S) is necessary.

Though the A X DR interaction does not support the potential observer theory, it does show evidence for an effect of future observers, at least in relationship to another variable.[1] The treatment of the analyzer factor as a random variable allows us to consider the two analyzers in the present study as a sample from the pool of potential analyzers; thus, the A x DR interaction can be generalized beyond the performance of these two specific individuals.

This result has serious implications for the "Edinburgh split" (Broughton & Millar, in press) in which a body of data is divided into two sets, one to serve as a "pilot" and the other a "confirmation" series. This method has the advantage of ensuring relatively consistent conditions between the series, particularly with respect to the experimenter's expectancies and their possible influence on subjects' scores. However, if analyzer differences can occur between two individuals (or the same individual at different times) then one must take into account the analyzer's expectancies before each of the two analyses, which are clearly not the same.

The difference in analyzer effects between the first and second halves of the session deserves some comment. This result is due primarily to a decline of analyzer effects at the 75o and 87.5% destruction rates (though an incline at the 0o destruction rate was also found). The decline of the analyzer differences between the first and second halves of the sessions, though not formally predicted, has some precedents in the literature. For example, Bierman and Houtkooper (in press) discovered analyzer differences to be localized in the first portion of subjects' runs. Feather and Brier (1968), in their two-series study of the checker effect, found in both series that this effect was very strong in the first runs. And, of course, decline effects have commonly occurred in more conventional tests of psi performance. The fact that this internal "sign of psi" is found for analyzer differences strengthens its claim as a genuine psi effect. In addition, the decline may prove to have practical consequences as well if the first half of a data set can be used to "drain off" analyzer effects. Such a concept needs to be tested empirically.

---

[1] An alternative hypothesis that the results are caused by the subject's contemporaneous psi acting in a goal-oriented manner so as to create this pattern in the data can be applied to any parapsychological outcome and is at present untestable. Until boundary conditions can be established for psi-mediated experimenter effects this problem will affect the validity of any experiment.

REFERENCES

BIERMAN, D. J., & HOUTKOOPER, J. M. The potential observer effect, or the mystery of irreproducibility. In W. G. Roll (Ed.), *Research in parapsychology,* 1979. Metuchen, NJ.: Scarecrow Press, in press.

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. In Joseph A. Steger (Ed.), *Readings in statistics for the behavioural scientist.* New York: Holt, Rinehart & Winston, 1971, 311-330. Originally published in *Psychological Bulletin, 57,* 1960, 49-64.

BROUGHTON, R., & MILLAR, B. Split analysis technique for robust effects. In W. C. Roll (Ed.), *Research in parapsychology,* 1979. Metuchen, N.J.: Scarecrow Press, in press.

FEATHER, S. R., & BRIER, R. The possible effect of the checker in precognition tests. *Journal of Parapsychology,* 1968, 32, 167—175.

HARTWELL,J. W. A bound for the observational theories of psi. *European Journal of Parapsychology,* 1977, 2(1), 19-28.

HOUTKOOPER, J. M. A comment on Schmidt's mathematical model of psi. *European Journal of Parapsychology,* 1977, 2(1), 15-18.

KEPPEL, G. *Design and analysis: A researcher's handbook.* Englewood Cliffs, NJ.: Prentice-Hall, I 973.

MILLAR, B ., & HARTWELL, J. Dealing with divergence. In W. C. Roll (Ed.), *Research in parapsychology,* 1978. Metuchen, NJ.: Scarecrow Press, 1979.

SCHMIDT, H. Toward a mathematical theory of psi. *Journal of the American Society for Psychical Research,* 1975, 69, 301-319.

SCHMIDT, H. A suggested method for detecting psi with subjects whose performance fluctuates from trial to *trial. Journal of the American Society for Psychical Research,* 1977, *71,* 19-31.

SCHMIDT, H. A remark on the "divergence problem." *European Journal of Parapsychology,* 1978, 2, 163-166.

WALKER, E. H. Foundations of paraphysical and parapsychological phenomena. In Laura Oteri (Ed.), *Quantum physics and parapsychology.* New York: Parapsychology Foundation, 1975.

WALKER, E. H. Comparison of some theoretical predictions of Schmidt's mathematical theory and Walker's quantum mechanical theory of psi. *Journal of Research in Psi Phenomena,* 1977, 2(1), 570.

WEINER, D.H., & BIERMAN, D. J. An observer effect in data analysis? In W. G. Roll (Ed.), *Research in parapsychology,* 1978. Metuchen, NJ.: Scarecrow Press, 1979.

WILSON, E. B., & HILFERTY, M. M. The distribution of chi-square. *Proceedings of the National Academy of Science,* 1931, *17,* 68689.

This paper is based on a report presented at the 1980 Southeastern Regional Parapsychological Association conference in Winter Park Florida February 15 and 16 1980. The authors wish to thank Drs. Charles Akers and Donald Burdick for their valuable advice regarding the statistical analysis.