Journal of Parapsychology, Vol. 53, September 1989

IN THE EYE OF THE BEHOLDER

FURTHER RESEARCH ON THE "CHECKER EFFECT"

By Debra H. Weiner and Nancy L. Zingrone

ABSTRACT: In 1968, Feather and Brier reported the "checker effect," which is the apparently psi-mediated effects associated with the person who scores ("checks") precognition data and which is consistent with the observational theories (OT) holding that those who observe the call-to-target match in an ESP experiment can influence the outcome of that match. In a previous two-series conceptual replication of the Feather-Brier research, we obtained results that largely confirmed their findings and supported the OT interpretation of the effect. The present paper continues this work with two new series.

Series III, an attempted replication of Series II, did not confirm our previous findings, but we found, post hoc, an interaction between run condition and observational context of checking; this interaction remained significant even when corrected for selection (p = .0021). Series IV was designed to simultaneously replicate Series II and test an OT hypothesis regarding the unexpected result in Series III. Series IV replicated Series II: outcomes were influenced by the checker, and this influence depended on the conditions under which call-to-target matches were observed. The hypothesis regarding Series III's interaction result could not be confirmed because the interaction did not recur.

In this paper, we consider the possibility of various artifacts and reject it. We argue that the results best support an observational model of the checker effect. We also discuss two trends (psi-missing in data not checked by the test administrator and significantly low variance in data checked under certain observational conditions) as well as implications for the role of observation.

In a previous paper (Weiner & Zingrone, 1986), we reported the results of a two-series conceptual replication of Feather and Brier's (1968) research on "the checker effect," which is based on the hypothesis that the person who "checks" precognition guesses against targets may play some role in how the results turn out, a role not explained by checking errors or other normal mechanisms. In that same report we showed how the checker effect is consistent with the "observational theories" (OT) of Schmidt (1975) and Walker (1975), and we presented results supporting the observational interpretation. Here we present a continuation of that work.

The authors wish to thank Ms. Vicki Newsome and Dr. Paul Brandes and their students, as well as the students of Sanderson High School gifted and talented program, for their participation. Thanks also go to Linda Ironside, Linda Vann, and Michelle Davis, who served as assistants during various portions of this work.

Background[1]

Feather and Brier (1968). Following laboratory anecdotes suggesting that results of precognition runs arbitrarily divided between two checkers differed significantly, Feather and Brier carried out the first experimental test of "the checker effect." One experimenter administered four-run precognition tests to subjects who were also asked to indicate which two runs they thought the test administrator would check and which two would be checked by an unnamed colleague (the other experimenter). Prior to being checked, runs were randomly divided between Feather and Brier. In their first series (FB-I), runs correctly predicted to be checked by Feather, the test administrator, had significantly higher scores than runs checked by Feather but predicted for the anonymous colleague. No such difference was found in Brier's data, however. In a confirmation series (FB-II), Feather and Brier switched roles. Again, runs correctly predicted to be checked by the test administrator had higher scores than those checked by him but predicted for the other checker. Again no such pattern occurred in data checked by the anonymous colleague.

The important feature of these results is that they depended not just on who subjects *thought* would check their runs but also on who actually *did* check them. As Feather and Brier put it, "Since the significance occurs only on those runs which the experimenter [test administrator] checked, it appears that the person who actually checks the test is having some effect upon the scores of the test he is checking" (Feather & Brier, 1968, p. 174). Further, a simple precognition hypothesis (that correctly predicting the checker—whether it be the test administrator or anonymous colleague—results in higher run scores) also does not explain their results. Because these data were independently double-checked (S. Feather, personal communication, July 19, 1984), this "checker effect" cannot be due to differences in checkers' accuracy or to motivated checking errors. Instead, it seems to be some sort of psi-mediated effect.

---

[1] For reasons explained in our earlier report, we will not discuss the Fisk-West research (Fisk & West, 1958; West & Fisk, 1953), in which checker differences and target-preparer differences are confounded. Similarly, the "checker" effect in the Houtkooper-Haraldsson study (1985) is, as the authors note, actually a second-order ("future") observer effect and is not directly germane to the present discussion. For the sake of clarity, the present review is restricted to studies in which the checker influence was deliberately tested and in which checkers were the first observers of the call-to-target matches. Although it is beyond the scope of this paper, this literature should be reviewed in conjunction with that of second-order observer effects (e.g., "analyzer effects") and the divergence problem (Schmidt, 1975) in general.

Subsequent checker-effect research. Prior to our work, there was only one attempted replication of the Feather-Brier experiments, a two-series study (Kennedy, O'Brien, O'Brien, & Kanthamani, 1977), which did not obtain significant results.[2] However, there are two other reports that shed light on the checker effect. One is Kennedy and Taddonio's (1976) reanalysis of the Feather-Brier data showing that runs checked by Brier when he was the anonymous checker were significantly below chance and significantly lower than Feather's when she played that role. Kennedy and Taddonio suggested that this result may reveal a psi-based experimenter effect.

The other relevant report is O'Brien's (1979) study of checker expectancy. In each of two series, O'Brien collected precognition data and divided the runs randomly between two checkers, one of whom had been led by O'Brien to expect above-chance scores and the other to expect below-chance scores. O'Brien independently double-checked all data to rule out checking errors. In the first series, the hitting-expectancy checker obtained above-chance scores ($CR = 2.32$, $p < .02$, two-tailed), whereas the missing-expectancy checker obtained chance scores, resulting in a suggestive difference (χ^2 [1] $= 3.77$, $p < .06$) in the direction of the induced expectation. A second series, with two new checkers, obtained scores significantly in the direction of the induced expectation (hitting-expectancy checker: $CR = 1.87$, $p < .05$, one-tailed; missing-expectancy checker: $CR = -2.44$, $p < .01$, one-tailed) and a significant difference between the two (t [30] $= 3.84$, $p < .001$, one-tailed). Thus, even though these checkers had had no contact with the subjects, what they believed about the outcomes of the runs they checked seemed to influence the results.

Interpretation of the checker effect. Three hypotheses about the mechanism of the checker effect were described in our earlier report. Briefly, they are as follows: (1) The subjects' precognition performance is affected by the subjects' unconscious reaction to information (obtained psychically) concerning who will check their guesses. (2) There is a psi-based experimenter effect (most likely nonintentional PK on random processes involved in determining targets) that results in a favorable outcome. (3) An observer effect occurs when checkers compare guesses to targets.

The rationale for the third hypothesis comes from the observational theories that hold that psi occurs when quantum mechanically

---

[2] Previously we incorrectly reported that this was a four-series study (Weiner & Zingrone, 1986).

random events are "measured," which, according to the theories' nonstandard interpretation of the measurement problem in quantum mechanics, occurs when the events are observed by conscious beings.[3] According to the OT, in an ESP test psi occurs not when guesses are made, as the traditional model holds, but when the targets are "measured" against the guesses, that is, when they are checked. This interpretation makes the checkers prime candidates for influencing the data they observe.

Effectively, the OT model of the Feather-Brier experiments inverts the usual order: instead of performing a mere clerical task, Feather and Brier are actually the unwitting subjects in a psi test that consists of two run conditions: runs predicted for themselves and runs predicted for their colleague. To understand Feather and Brier's specific finding (i.e., that there was a significant difference between correctly and incorrectly predicted runs only in the data of the test administrator) within this framework, we drew on O'Brien's work and on information from Feather and Brier regarding their expectations at the time of their research. Feather and Brier had not been blind to subjects' predictions of the checker (the run's "prediction category") as they checked their data. Further, both experimenters had expected that the test administrator would be more likely to obtain significant results (R. Brier, personal communication, November 25, 1985; S. R. Feather, personal communication, July 19, 1984). We hypothesized that the test administrator may have expected high scores on runs predicted for him/her and low or chance scores on runs predicted for the anonymous colleague and that, from O'Brien's work, such a difference in expectation when checking these runs could have produced a significant difference in the test administrator's data. The anonymous colleague, on the other hand, did not expect to obtain significant results and therefore approached the runs of the two prediction categories in what may have been a psychologically equivalent manner. Three of the four experiments in the present series addressed this hypothesis.

Series I and II. Our earlier work was a two-series conceptual replication of Feather and Brier's experiment (Weiner & Zingrone, 1986). In both series N.L.Z. was the test administrator and D.H.W. the anonymous checker. In Series I, students of a technical school's psychology courses were given four precognition runs and asked to guess which two runs would be checked by N.L.Z. These runs were

---

[3] We do not intend to discuss here the various controversies regarding the OT (e.g., Bierman, Houtkooper, & Millar, 1981; Braude, 1979, 1988; Millar, 1988; Phillips, 1984; Walker, 1984).

later divided between N.L.Z. and D.H.W. such that each checker received from each subject one run predicted for N.L.Z. and one run predicted for the anonymous colleague, in a counterbalanced order. D.H.W., blind to the subjects' calls, obtained an entry point into a computer file of random numbers (RAND Corporation, 1955) and produced computer-printed target sequences. Like Feather and Brier, we were not blind to the subjects' predictions of the checker while we checked the data.

We formed two hypotheses. From Feather and Brier's results, we predicted that the scoring difference between prediction categories would vary depending on who actually checked the runs, leading to a checker-by-prediction (C × P) interaction. This hypothesis was not confirmed, though it received indirect support from a significant difference (t [14] = 2.42, p = .03, two-tailed) between prediction categories for one checker (D.H.W.) but not for the other. The second hypothesis, based on Kennedy and Taddonio's reanalysis of the Feather-Brier work, was that a difference in checkers' scoring rates (checker main effect) would be obtained. This prediction was confirmed (F [1,59] = 8.14, p = .006). N.L.Z.'s scores were suggestively above chance ($p < .061$, two-tailed) whereas those of the anonymous colleague, like Brier's when he played that role, were significantly below chance (t [29] = − 2.13, $p < .05$, two-tailed).

In Series II, the data of 52 university and intellectually gifted high-school students in the Chicago area were used to retest Series I's predictions and to explore the hypothesized role of checker expectation. That hypothesis rested on the assumption that knowing for whom the run had been predicted led Feather and Brier to form certain expectations, which then influenced the outcomes of the runs they checked. Because we acted as both checkers and experimenters, we could not test this supposition directly by manipulating checker expectation as O'Brien had, but we could manipulate whether we knew for whom the run had been predicted, information that was necessary for such expectations to arise. Therefore, on half the runs of Series II we were blind to subjects' predictions while we checked the data. This manipulation also served more generally to test the observational model of ESP: by positing that observation of call-to-target matches is the crucial action in a standard ESP test, the observational model of ESP can be tested by systematically altering some aspect of that observational process (e.g., Broughton, 1977). The blindness manipulation altered the informational context (and thereby any of the attendant motivational, emotional, or cognitive contexts) of the call-to-target observation.

The results of Series II supported the observational model because the checkers' impact on differences between run conditions depended on the observational context of checking the runs (C × P × B interaction: F [1, 200] = 3.96, $p < .05$). When checkers were blind to the run's prediction category, there was absolutely no evidence of checker influence. Conversely, when checkers were *not* blind to subjects' predictions of the checker, both experimental predictions were confirmed. Further, the results in the not-blind condition were similar to those of Series I and the original Feather-Brier work, all of which were likewise checked under not-blind conditions. In the not-blind data of Series II, there were once again the following two results: (a) a significant difference in checkers' scoring rates (F [1, 200] = 5.86, $p < .025$), with D.H.W. again obtaining below-chance scores (t [51] = − 2.00, $p < .051$, two-tailed) and N.L.Z. obtaining above-chance scores, and (b) a significant difference between prediction categories in D.H.W.'s not-blind runs but not in N.L.Z.'s, this time resulting in a significant C × P interaction (F [1, 200] = 7.93, $p = .01$). Pooling across blind and not-blind conditions revealed a significant overall C × P interaction (F [1, 200] = 3.96, $p < .05$) and a suggestive difference in checkers' scoring rates (F [1, 200] = 2.69, $p = .10$).

These results support the observational interpretation of the checker effect in that they show that a manipulation of the checkers during the observation of call-to-target matches affected the overall outcome. This interpretation was further supported anecdotally by our later discovery that D.H.W., who in both series obtained significant differences between prediction categories, had been very aware of these categories during checking, whereas N.L.Z., who obtained nearly equal scores in the two categories, had paid relatively little attention to this information. This discovery, although post hoc, nevertheless fits well with the results of our planned analyses in that it implies that checker awareness of relevant information during the checking process plays a role in the outcome. The observational interpretation, however, was contradicted by a secondary analysis intended to replicate a post hoc Feather-Brier finding (i.e., that the checker effect was concentrated in the subjects' Run 1's). Run order and checker observation order had been confounded in the Feather-Brier work but not in ours. Analyses testing the C × P interaction against run order and checker observation order found that the C × P interaction was concentrated in runs that subjects completed relatively early in the test and was not related to the order in which the checkers observed the runs. This relationship of the "checker" effect to a condition of testing (run order) suggested that the subjects also played a role in the outcome.

Implications. Whether the subjects or the checkers (or both) were the psi sources in Series II, the results indicate that the checkers' knowledge of subjects' predictions was important. In our previous report we described two possible reasons for this: (1) knowledge of subjects' predictions made the observation of the call-to-target match *meaningful,* and (2) this knowledge made the observation *complete.*

By "meaningful observation" we meant that with this information the checker would know how the run score would fit within the overall goal of confirming the experimental predictions, though other aspects of meaningfulness were not excluded from consideration.[4] By "complete" observation we meant that the observation provided complete reduction of uncertainty. Using as a model von Lucadou and Kornwach's systems-theoretical approach to psi (Kornwachs & von Lucadou, 1979; von Lucadou & Kornwachs, 1977a, 1977b, 1980a, 1980b, 1983), we considered Series II as being a "system" that changed over time. Tracing the "observational histories" of data as they were "processed" through the system—from checking to the "final output," the analysis of variance (ANOVA)—we find that the blind and not-blind runs took somewhat different paths. For each run, four items of information had to be obtained to determine the composition of the ANOVA cells and, therefore, the results: (1) the identity of the checker; (2) the condition of observation (blind or not-blind); (3) whether the run was predicted for N.L.Z.; and (4) the run score. During checking, only the last three items can vary because the identity of the checker is fixed. For the not-blind data, all three items of information were observed at approximately the same time and during the same procedure: checking the run. However, for the blind data, two items were observed during checking (observational condition, run score); the remaining item (prediction category) was obtained at a later time. Thus, for not-blind runs, complete reduction of uncertainty occurred in *one* step, whereas for blind runs it occurred in *two* steps. This separation, or possible differences in the psychological or observational conditions of the two steps, may explain why no checker effects were found in the blind data.

---

[4] As explained in our previous paper, the concept of a meaningful context of observation is *not* the same thing as von Lucadou and Kornwachs's concept of meaningful observation (von Lucadou & Kornwachs, 1980a). By their definition, the checkers' observation under blind conditions is meaningful as well.

SERIES III

Series III was a replication of Series II, with two deliberate changes in procedure: (1) like Feather and Brier, we switched roles, with D.H.W. administering the tests and N.L.Z. being the anonymous colleague, and (2) we made certain that on not-blind runs both checkers were aware of the subjects' prediction of the checker. All testing for both Series III and IV was completed before targets for Series III were generated.

Method

Subjects

In late October, 1985, D.H.W. gave lectures on parapsychology to an undergraduate communications class (Psychology Department, University of North Carolina at Chapel Hill) and to a special class of gifted and talented high-school students (Sanderson High School, Raleigh, NC). This population was similar to that of Series II in that it consisted of a combination of university and intellectually gifted high-school students, though they differed in geographical region and presumably in other, unmeasured, ways as well.

Students were invited to participate in the experiment to be given at the end of the lecture. As in Series I and II, students were told that they could choose not to participate (by simply not completing the tests) or to take the precognition test but have the scores excluded from the experiment (by completing the test but not signing the consent form). We used these measures to minimize any perceived pressure to participate and to thereby increase genuine voluntarism.

As in Series II, data were used from only those subjects (a) who were students 18 years of age or older with a signed consent form (or students under 18 years of age with a consent form signed by a parent or guardian); (b) who completed all 25 trials of each run; and (c) who correctly followed instructions to indicate two runs to be checked by D.H.W. Of the 46 students tested, 40 (87%) (30 females, 10 males) met these criteria. The subjects ranged in age from 16 to 22 (mean = 20).

Procedure

The same test procedure used in Series I and II was used in this study. At the end of her lecture, D.H.W. described the test and the students' participation options and administered four runs of a standard 25-trial precognition test using ESP-card symbols. Each run was printed on a separate sheet for easy distribution to checkers. The participant was asked to guess which two runs would be checked at a later date by D.H.W. and which two by an unnamed colleague, and to indicate this prediction by marking either "yes" or "no" in answer to a question printed at the top of each call sheet ("Do you think this run will be checked by the Experimenter?").

Consent forms were included with the call sheets. (Consent forms for high-school students had been distributed earlier to obtain parental signatures.) Subjects were given the opportunity to request a report of their scores, if they wished.

Preparation of Call Sheets

Remaining blind to the subjects' calls, D.H.W. supervised an assistant who prepared the call sheets for checking. The procedure used in Series II was used here, with minor modifications to facilitate double-checking. (See Weiner & Zingrone, 1986, for details.) An assistant alphabetized by surname the call sheets of eligible participants and separated the runs predicted to be checked by D.H.W. (runs with "yes" in answer to the "experimenter as checker" question) from those predicted for the anonymous checker (runs with "no" in answer to the question). Alphabetical order of the subjects' data was maintained throughout the preparation. Using photocopies of the call sheets (originals were stored in a locked cabinet inaccessible to either experimenter), she divided the yes runs into two sets, one for each checker, in an ABBA order (D.H.W., N.L.Z., NL.Z., D.H.W.), did the same for the no runs, and interfiled each checker's yes and no runs so that the two runs from a given subject were together. The assistant then divided D.H.W.'s call sheets into blind and not-blind runs by designating the *pair* of runs from a subject as "not blind" or "blind" in an ABBA order (not-blind, blind, blind, not-blind). Taking the not-blind pairs as a group, the assistant alternated the yes and no runs in an ABBA order (yes, no, no, yes). The blind pairs were arranged according to a closed-deck random sequence from the PDP 11/20 RAN function so that half of them began with a yes run.[5] Checkers thus could not infer the prediction category of a blind run from the order of presentation.

---

[5] This order was as follows, with A indicating a run predicted for D.H.W. and B, a run predicted for the other checker: 1(AB), 1(BA), 3(AB), 2(BA), 1(AB), 3(BA),

After the assistant recorded onto data sheets information to help the double-checkers reassemble call sheets, she cut the blind sheets in half so that checker prediction was not available. Blind and not-blind sets were then interfiled back into the original alphabetical order and numbered 1 to 80 for presentation to the checkers. At no time during this process did either checker have access to the research materials.

In summary, the pile of call sheets received by each checker began: not-blind/yes; not-blind/no; blind/? [yes]; blind/? [no]; blind/? [no]; blind/? [yes]; not-blind/no; not-blind/yes; and so on. Checker, blindness, and prediction category were manipulated orthogonally, and the number of runs in all possible combinations of conditions was the same. All four runs from a given subject were in the same blind or not-blind condition.

Target Generation

N.L.Z. generated target sequences using the same procedure as in Series I and II. (See Weiner & Zingrone, 1986, for details.) Blind to subjects' calls, N.L.Z. obtained an entry point into a computer file containing the RAND table of a million random digits (RAND Corporation, 1955). The entry point procedure was a modification of the standard method described by Morris (1968), which presumably was the procedure used by Feather and Brier. A 10-sided die was rolled eight times to produce four two-digit numbers. These were multiplied together; the product was then multiplied by itself backwards and the square root of that product was taken. The five most significant digits were taken as the entry point. To put the entry point within range of the line numbers of the random number table, the leftmost digit was changed to a 0 if it was even and to a 1 if it was odd. This process was witnessed and double-checked by a colleague, who retained a copy of the entry point for security purposes.

After the appropriate section of the random number file was accessed, a computer program translated the digits into ESP symbols using a standard code (Rhine & Pratt, 1957, p. 151): 1,6 = "O"; 2,7 = "+"; 3,8 = "="; 4,9 = "L"; 5,0 = "*". The program printed target sequences separately for the two checkers and stored coded sequences into separate computer files for double-checking. In ac-

---

1(AB), 1(BA), 2(AB), 2(BA), 1(AB), 1(BA), 1(AB). (The letters refer to the order of runs within the pair; the numbers refer to the number of consecutive pairs with that order.)

cordance with the Feather-Brier procedure, all of the test administrator's targets followed the entry point, with targets for the anonymous colleague following immediately after them. A separate target order was created for each run.

Data Checking

Once the targets and call sheets were ready, D.H.W. received both sets of call sheets (hers and N.L.Z.'s) from the assistant. She gave N.L.Z. her set and at the same time received from N.L.Z. her own target sheets. (By this type of exchange, neither checker had prior access to both call sheets and targets at the same time.) Upon receiving these materials, we proceeded to check the data. To make ourselves equally aware of the run's prediction category, we agreed to circle the subject's response (yes or no) to the prediction question prior to checking the run. Furthermore, it was planned that N.L.Z. would pattern her checking strategy more closely to D.H.W.'s, which involved a playful, gamelike attitude toward the task. As before, the checkers agreed to check the runs in the order given and not to discuss their results until both parties had completed the task.

Double-checking, Decoding, and Feedback

Using the data record sheets prepared by the first assistant, a second assistant assembled unmarked photocopies of the original call sheets to match the order used by the checkers. Using these sheets, two assistants, one working with D.H.W.'s data and another with N.L.Z.'s data, entered calls into computer files and computer-scored them against the checkers' target files. Discrepancies between computer- and hand-checked scores were noted by the double-checker and resolved by the checker working with her assistant. Double-checking uncovered errors in 2.9% of D.H.W.'s checking and in 0.45% of N.L.Z.'s checking. This represents a high percentage (97.1% and 99.55%, respectively) of trials correctly observed by the checkers, supporting the assumption that the checkers were the first observers of the correct call-to-target matches. (If this percentage had been low, one could argue that the double-checkers, as the first observers of correct matches on a substantial proportion of trials, should be considered as potential influencers of the outcomes.)

In August, 1986, all persons who had requested feedback, including those who did not meet the selection criteria, received a letter giving their scores, information for evaluating them, and a brief description of the experiment's purpose and preliminary findings. Targets for unusable runs were generated, for feedback purposes, after analysis of study data. Thirty-nine subjects (97.5%) requested feedback.

Experimental Predictions and Method of Analysis

It was planned to analyze the data by a three-way, fixed-effects ANOVA, with checker, prediction category, and blindness as factors. On the basis of our previous findings, the following predictions were made:

1. *There would be a significant "checker effect" (C × P interaction) in the not-blind data but not in the blind data.* This prediction was tested by the C × P × B interaction term. We expected a significant simple interaction in the not-blind data and a nonsignificant simple interaction in the blind data.

2. *Checkers' scoring rates would differ significantly only in data checked under not-blind conditions.* This prediction was tested by the checker main effect in the not-blind data.

Our prediction of a C × P interaction in the not-blind data was not limited to a significant difference between yes and no runs for one checker only. It will be recalled that in our previous work the checker who obtained significant differences between yes and no runs was the one who had paid the most attention to this information. Because in the present series we had taken steps to ensure that both checkers were aware of the subjects' predictions during checking, we did not necessarily expect this exact pattern to replicate. (In fact, it would be more in line with the observational interpretation if it did *not*.) Also, we did not prespecify the direction (e.g., yes > no) of differences between prediction categories. Similarly, because of the reversal of roles, we did not prespecify whether one checker would obtain independently significant results or whether these scores would be above or below chance.

It was planned that if either prediction was confirmed, we would test for effects of subject run order and checker observation order. Additional secondary analyses would attempt to confirm Series II's significant overall C × P interaction and suggestive checker main effect (pooling across blind and not-blind runs).

All analyses were double-checked by assistants who independently collated scores, entered them into computer files, and computer-compared analysis and double-check files.

TABLE 1
RESULTS OF SERIES III

Source	*SS*	*df*	*MS*	*F*
Checker	1.60	1	1.60	< 1
Prediction	1.22	1	1.22	< 1
Blindness	4.90	1	4.90	1.32
C × P	0.22	1	0.22	< 1
C × B	0.40	1	0.40	< 1
P × B	50.62	1	50.62	13.59*
C × P × B	4.22	1	4.22	1.13
Error	566.30	152	3.73	

*p = .0003 (not predicted); corrected for selection: p = .0021.

Results

Planned Analyses

Neither experimental prediction was confirmed (see Table 1). The secondary predictions were likewise not confirmed. Because there was no evidence of checker influence, there was no basis for testing effects of subject run order and checker observation order (but see the post hoc analyses that follow).

Post Hoc Analyses

An unexpected outcome of the ANOVA was a significant prediction-by-blindness (P × B) interaction (F [1, 152] = 13.59), which remained statistically significant when corrected for selection out of seven nonsignificant terms (p = .0021). When checked under not-blind conditions, runs predicted for D.H.W. were higher than runs predicted for the anonymous checker (F [1, 76] = 5.03), with hitting in the former (M = 5.43) and missing in the latter (M = 4.47). When checked under blind conditions, this difference was in the opposite direction (F [1, 76] = 8.75), with missing on runs predicted for D.H.W. (M = 4.65) and hitting on runs predicted for the anonymous checker (M = 5.95, t [39] = 2.88). These differences were larger in D.H.W.'s data (for blind data, t [38] = 3.16; for not-blind data, t [38] = 1.85). The differences in N.L.Z.'s data were in the same direction as D.H.W.'s but were smaller in magnitude.

In the spirit of the secondary analysis described above, we carried out a post hoc analysis (ANOVA) to see whether the P × B interaction was related to subjects' run order or checker observation order. Both the P × B × Run Order and P × B × Observation Order interaction terms were nonsignificant.

Discussion

Clearly, Series III did not confirm its predictions. It did, however, obtain a post hoc interaction between condition of checking (blind vs. not-blind) and prediction category. In comparing the procedures of Series II and III, we discovered one potentially important difference: In Series II, the second observational step—decoding the blind runs into those predicted for the test administrator and those for the anonymous colleague—was performed by an assistant, whereas in Series III it was done by the two checkers working together. Could the fact that the parties who performed the decoding ("completing" the observation) of blind runs were knowledgeable about the experimental predictions and interested in success have allowed them unwittingly to influence the data at that stage? Series IV tested this speculation by the use of two blind conditions, one to be decoded by an assistant (as in Series II) and the other to be decoded by the checkers (as in Series III). In this way, both Series II and III could be replicated simultaneously.

SERIES IV

Subjects

The subjects were psychology students at Durham Technical Institute (Durham, NC) who were tested by D.H.W. following her lectures on parapsychology to their classes. The test procedures were the same as those used in Series III. One class was tested in early May, one in late August, and two in mid-November 1985. The subjects ranged in age from 18 to 60 (Mdn = 23); 23 were female. Of 45 students, 33 (73%) met the selection criteria.

Method

Preparation of Call Sheets

Remaining blind to subjects' calls, D.H.W. supervised an assistant who prepared the call sheets for checking. We used the same procedure as in our previous work, modified to allow for two blind conditions. The assistant followed the same steps used in Series III until the point at which she was to divide each checker's call sheets into blind and not-blind conditions. In this series she divided the data into three groups—not-blind, "Blind #1" (B1), and "Blind #2" (B2)—according to a closed-deck random order from the PDP 11/20 RAN function. The yes and no runs in the not-blind data were arranged in ABBA order, as in our previous work. As before, the assistant generated a closed-deck random sequence from the PDP 11/20 RAN function to determine the order of the B1 yes/no pairs such that 50% of them began with the run predicted for D.H.W. The yes/no pairs in the B2 data were treated in the same way except that a different closed-deck sequence was obtained and the decoding information was kept on a separate record sheet.[6]

After the assistant recorded various items of information to facilitate double-checking, she cut the B1 and B2 sheets in half so that checker prediction was not available. Not-blind, B1, and B2 runs were then interfiled with the original alphabetical order and numbered 1 to 66 for presentation to checkers.

In summary, each checker received a pile of 66 call sheets. For 22 sheets the checker could see for whom the run had been predicted, whereas for the other 44 sheets this information was not available. Half of the 44 blind sheets belonged to the B1 condition and half to the B2 condition, but checkers could not distinguish these conditions. The blind sheets (both conditions) and not-blind sheets were intermixed in a random order.

Target Generation

N.L.Z. generated target sequences using the procedures previously described. The entry-point procedure was again witnessed by a colleague for security purposes.

Data Checking

Target lists and call sheets were exchanged as in Series III. As before, checkers agreed to check the runs in the order given and

[6] This order was as follows, with A indicating a run predicted for D.H.W. and B, a run predicted for the other checker: B1: 2(AB), 1(BA), 2(AB), 2(BA), 1(AB), 2(BA), 1(AB). B2: 2(BA), 1(AB), 3(BA), 2(AB), 1(BA), 2(AB). (The letters refer to the order of runs within the pair; the numbers refer to the number of consecutive pairs with that order.)

not to discuss the results with anyone prior to both parties finishing the task. In addition, N.L.Z. stated in advance that she would attempt to maximize the difference between prediction categories in her not-blind runs in her favor (i.e., no runs > yes runs).

Double-checking, Decoding, and Feedback

The same procedures for double-checking previously described were used in this series. Double-checking uncovered errors in 1.5% of D.H.W.'s checking and 2.2% in N.L.Z.'s checking, which supports our assumption that the checkers are the first observers of the (correct) call-to-target match.

The experimenters decoded their own B1 data, using the record sheets prepared by the assistant. The decoding list for the B2 condition was given to another assistant, a new employee blind to the purpose of the experiment, who was asked to compile yes and no responses and run scores for each experimenter. (As a manipulation check, a written, open-ended question asked her about her thoughts during the task and her understanding of the study. Her response indicated that she was not aware of the purpose of the experiment.)

In February, 1987, all persons who had requested feedback, including those who did not meet the selection criteria, received a letter giving their scores, information for evaluating them, and a brief description of the experiment's purpose and preliminary findings. Targets for students who had not met the selection criteria were generated after analysis of the study data. All subjects requested feedback.

Predictions

Primary predictions were that Series IV would replicate the two significant results of Series II. Secondary predictions tested the role of decoding in the P × B interaction (from Series III) or called for replication of secondary or suggestive results in Series II. We planned to conduct exploratory analyses of subject run order and checker observation order if significant evidence of checker influence was obtained. All predictions were stated in advance of checking the data.

Primary Predictions

1. *There will be a checker effect (C × P interaction) in the not-blind data only.* Thus, it was predicted that there would be a significant simple

C × P interaction in the not-blind data and a nonsignificant interaction in the blind data, leading to a C × P × B interaction.

2. *There will be a checker main effect in the not-blind data only.*

Secondary Predictions

1. *There will be a P × B interaction in the 2 × 2 × 3 ANOVA, with significant differences in prediction category in the not-blind and B1 levels and no significant difference in the B2 level.* This prediction tests the "decoding hypothesis" about Series III's P × B interaction.

2. *There will be overall evidence of a checker effect (C × P interaction).* This prediction follows from Series II and was to be tested both (a) overall (not-blind, B1, and B2 data) and (b) in those data comparable to Series II (not-blind and B2 data only).

3. *There will be an overall difference in checkers' scoring rates.* This prediction also follows from Series II and was to be tested by the checker main effect. It, too, was tested both (a) overall and (b) in the data comparable to Series II.

4. It was planned that in her not-blind data N.L.Z. would attempt to produce greater scores in the no than in the yes runs and that D.H.W. would attempt to do the opposite. Given the difficulty of producing directional results in parapsychology, these differences were prespecified as two-tailed tests.

Method of Analysis

The planned method of analysis was a 2 × 2 × 3 (checker by prediction by blindness) fixed-effects ANOVA. Because both primary predictions were based on the results of Series II, we planned to test them using the B2 data, the blind condition comparable to that of Series II. To avoid extraneous effects from the B1 condition, we tested the primary experimental predictions by comparing the mean square of the appropriate term (e.g., C × P interaction in not-blind data) against the within-cell error from the 2 × 2 × 3 ANOVA. (This error term was chosen because, being based on all cells, it makes maximum use of the data and serves as the best estimate of experimental error.) This approach was also used to test secondary predictions 2b and 3b; the remaining predictions were based on the full 2 × 2 × 3 ANOVA.

TABLE 2
RESULTS OF PRIMARY PREDICTIONS, SERIES IV

Source	*SS*	*df*	*MS*	*F*
Not blind:				
Checker	5.11	1	5.11	1.26
Prediction	0.02	1	0.02	< 1
C × P	21.84	1	21.84	5.41*
Blind (B2):				
Checker	11.00	1	11.00	2.72
Prediction	5.82	1	5.82	1.44
C × P	0	1	0	0
C × P × B	10.92	1	10.92	2.70
Error	484.54	120	4.04	

*$p < .025$.

Results

Primary Predictions

Prediction 1. As predicted, the C × P interaction was significant in the not-blind data (F [1, 120] = 5.41, $p < .025$) and was not significant in the blind (B2) data (F [1, 120] = 0.00);[7] however, the C × P × B interaction was not significant (F [1, 120] = 2.70, p = .103). (See Table 2.) As she had attempted to do, N.L.Z. obtained higher scores in her not-blind no than her yes runs (t [20] = 1.88), marginally significant by a two-tailed test (p = .075, two-tailed). D.H.W. obtained nonsignificantly higher scores in her not-blind yes vs. no runs (t [20] = 1.32).

Prediction 2. The checker main effect in the not-blind data was not significant (F [1, 120] = 1.26).

Secondary Predictions

Prediction 1. The P × B interaction was not significant (F [1, 120] < 1).

Prediction 2. The C × P interaction was not significant either overall (F [1, 120] < 1) or in the data comparable to Series II (F [1, 120] = 2.70, p = .103).

[7] For completeness, we report that the C × P interaction in the B1 condition was also nonsignificant: F (1, 120) < 1.

Prediction 3. The overall checker main effect was nonsignificant (F [1, 120] = 2.30, p = .13). The checker main effect in the data comparable to Series II was suggestive (F [1, 120] = 3.85, p = .052).

Exploratory analyses of run order and checker observation order were precluded by very small cell sizes.

Summary

The results of Series IV show remarkable similarity to the study we were attempting to replicate (Fig. 1). The simple interaction in the not-blind condition is significant, as predicted, and the simple interaction in the blind condition is not. As in Series II, this latter interaction term is extremely small (F = 0). The hypothesized three-way interaction was not confirmed, however, possibly because of the low power of the test (cell n's = 11). There was no evidence of a difference in scoring rates between checkers in the not-blind runs, though there was a suggestive difference (p = .052) across data comparable to Series II, with (post hoc) overall psi-missing in data checked by the anonymous colleague (t [65] = − 2.25). We could not evaluate whether Series III's P × B interaction was related to differences during the second step of observation because the P × B interaction did not recur.

Discussion

Because our primary finding—a significant difference between run conditions when checkers knew the run condition—is consistent with hypotheses of conscious or unconscious motivated checking errors and experimenter fraud, the first issue to address is whether these results can be explained as artifacts. This question was discussed in our previous report, and because our methodology here is essentially the same, those arguments apply to the present work as well. Stated briefly, they are as follows: (1) The criticism of inadequate target randomness has been eliminated by the use of well-tested random digit tables and our system of dividing runs among conditions, which effectively scattered the blocks of digits entering into any one condition in a fixed but arbitrary way across the set of digits used in this study. (2) Checking errors were caught and corrected by blind, independent double-checking. (3) The argument of experimenter fraud has been countered by (a) the strict division of labor between D.H.W. and N.L.Z., which precluded either of them from access to both calls and targets prior to

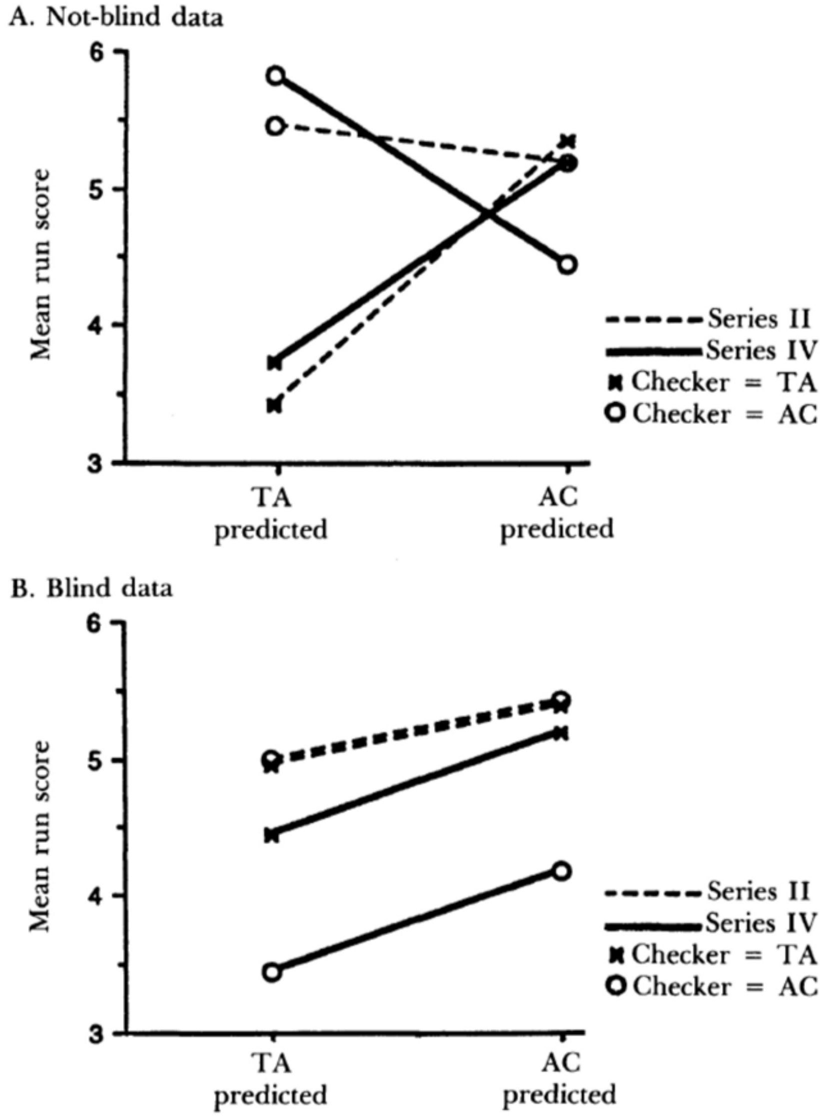

Figure 1. Results of Series II and IV showing the impact of the informational context (blind vs. not-blind) on the "checker effect" (C × P interaction). For consistency between series, results are graphed in terms of the experimenters' roles (test administrator [TA] vs. anonymous colleague [AC]). For both graphs the mean chance expected (MCE) run score is 5.0.

checking; (b) the use of published random sequences and a fixed method for determining the entry point, with a copy of the entry point secured with a colleague prior to target generation; (c) the use of independent assistants to prepare call sheets; (d) a fixed system of arranging call sheets (i.e., alphabetical and ABBA orders), which prohibited checkers from rearranging them to obtain better results; and (e) storage of original call sheets out of reach of both checkers. The replication in Series IV of Series II's results (which themselves replicated Series I and conceptually replicated the original Feather-Brier work) strengthens confidence that the results are not statistical noise. Therefore, we conclude that the present results cannot be explained on the basis of these artifacts.

Overview of the Checker-Effect Literature

To date there have been 10 experimental studies,[8] composed of 14 data sets, testing for checkers' influence on psi test scores. These have looked either at an interaction between the checker and some comparison of interest (e.g., yes vs. no runs) or at simple differences in checkers' scoring rates. The results of these studies are summarized in Table 3. As can be seen, 33% of the data sets testing for checker influence of the interactional type and 31% of those testing for differences in checkers' scoring rates are significant, compared to the 5% expected under the null hypothesis. From OT considerations, we would not necessarily expect checker effects to remain the same when observational conditions are altered. Viewing this literature separately for blind and not-blind conditions, then, we see that none of the data sets checked under blind conditions show checker influence, whereas 50% of the not-blind data sets show checker influence of the interactional type and 44% show significant differences in checkers' scoring rates. Although the number of studies in this literature is still extremely small and the "file-drawer" problem is not taken into account here, these percentages encourage continued investigation of checker influence.

The Mechanism for the Checker Effect

In our previous report we discussed our findings in terms of three proposed mechanisms for the checker effect: (1) subjects' psi-mediated

[8] Since we carried out the present work, an additional checker-effect study has been published (Kreiman, Ivnisky, & Marquez, 1987). Although the authors consider their work a conceptual replication of ours, their study had substantial differences in its design (single subject; computer precognition test; the subject designated rather than guessed some runs for checkers; and, most important, checker "knowledge" was redefined as whether the checkers knew the tests were being conducted rather than their knowing the run's prediction category).

TABLE 3
OVERVIEW OF CHECKER-EFFECT STUDIES

Series	Significant interaction type?[a]	Significant checker difference?
FB-I (1968)	yes	yes[b]
FB-II (1968)	yes	
Kennedy I (1977)	no	no
Kennedy II (1977)	no	no
O'Brien I (1979)	—	no
O'Brien II (1979)	—	yes
WZ-I (1986)	no	yes
WZ-II/NB (1986)	yes	yes
WZ-II/B (1986)	no	no
WZ-III/NB	no	no
WZ-III/B	no	no
WZ-IV/NB	yes	no
WZ-IV/B1	no	no
WZ-IV/B2	no	no
N significant studies	4/12 (33%)	4/13 (31%)
N significant studies (not-blind only)	4/8 (50%)	4/9 (44%)

[a]"Interaction type" refers to checker influence on a comparison between conditions (e.g., yes vs. no runs).
[b]Kennedy and Taddonio's (1976) reanalysis compared data between FB series.

reaction to the checker; (2) experimenter psi (PK) on random processes entering into target generation; and (3) observer effects during the checking process. The Intuitive Data Sorting (IDS) model (May et al., 1985), not explicitly mentioned in the previous report, is in this case a variant of the experimenter psi model because it, too, would hold that the experimenter used psi (precognition of the optimal time to generate an entry point) to take advantage of local biases in the random number table.

The subject-psi model is not supported by the primary finding that checker influence is restricted to runs checked under not-blind conditions: the subjects' reaction to the checker would not be expected to vary depending on what the checker knows or doesn't know while checking runs. In our design, an indirect causal route (subjects precognize the checkers' experience of checking, which for some reason influences their guessing accuracy) is experimentally indistinguishable from a direct route (checkers are psi sources influenced by their own experience) but is less parsimonious. This is not to say that subjects played no role in the experimental outcome: Series II found that the checker effect was concentrated in runs that subjects completed relatively early in the test. This finding, however, was not replicated in Series III and could not be tested in Series IV. At present, the best statement regarding this model is that the subject may play an interactive role in producing "checker" effects but is not likely to be solely responsible for such effects.

Hypotheses of psi-based experimenter effects do not apply well to our results for the following reasons: (1) Stanford (1981) has argued on theoretical grounds that the opportunity for psi-based experimenter effects should increase with the number of random decisions or events in which the experimenter participates; in our work virtually all major decisions (i.e., entry criteria, the entry-point algorithm, and arrangement of call sheets) were prespecified and fixed. (2) Morris's (1968) test of experimenter psi on an entry-point procedure (the same procedure used in our work) gave the best results in relatively short sequences (100 digits) after the entry point, which follows logically from the definition of a good random number table (no large-scale biases are present). The IDS model likewise finds a dependence of scoring on sequence length (May et al., 1985). In our work, only one entry point per series was used, generating 1,500, 5,200, 4,000, and 3,300 targets, respectively. (3) Most important, however, to explain our results as psi-based experimenter effects on the entry-point process, one must make ad hoc assumptions about the structure of the RAND table. Specifically, one must assume that the regions of localized nonrandomness tapped by experimenter psi exist in precisely those blocks of 25 digits, scattered throughout the digits used in our study, that created the targets of the not-blind condition. This scatter of digits entering into significant runs renders the application of the IDS model to the present work particularly difficult because a single decision (the process of obtaining one entry point) would have to be responsible for "sorting" a variety of sequences into opposite "bins." Further, one must assume that localized nonrandomness being tapped by experimenter psi is not of a simple sort (e.g., excess of even digits) but that it matched (and, for psi-missing runs, mismatched) subjects' idiosyncratic call biases. Even though complex random-number-table biases are theoretically possible, it is unparsimonious to assume they exist without supporting evidence.

The observer model of the checker effect is supported by the fact that manipulating the informational context of call-to-target match observations—a condition that, under traditional interpretations of precognition, would be irrelevant to scoring—appears to influence the outcomes of those matches. The observer model is further supported anecdotally by the relationship across the four series between N.L.Z.'s checking behavior on not-blind runs and her results. In Series I and II, she paid relatively little attention, while checking, to whether the run had been predicted for her, and her scores were similar between the two conditions (Series I: z = .354 vs. .355; Series II: M = 5.46 vs. 5.19). In Series III, she paid more attention to the run's prediction category and used the playful strategy D.H.W. had used; here the difference in her scores was greater and paralleled D.H.W.'s, but was not significant (M = 4.45 vs. 5.15). In the final series she again took a playful approach but this time made a special effort to enhance scoring in favor of runs predicted for her, which occurred (M = 5.18 vs. 3.73, p = .075, two-tailed). This pattern, though anecdotal, adds to the impression that the checker's psychological approach to, and motivations and intentions regarding, the checking task may influence the outcome.

Psi-Missing and the "Other Checker"

There is an interesting trend in the successful checker-effect studies for the test administrator to obtain scores above chance and the anonymous colleague to obtain scores below chance, sometimes significantly so. Elements of this pattern are observed in the first Feather-Brier series and in three of our own, with a significant difference in overall scoring rates in two studies (FB, WZ-I) and a suggestive difference in a third (WZ-II). In one case (WZ-I) the test administrator obtained suggestive hitting ($p < .061$, two-tailed); significant missing in data checked by the anonymous colleague was found in three studies (FB, WZ-I, WZ-IV). The pattern becomes sharper if we consider only data checked under not-blind conditions: in these data, three studies obtained significant between-checker differences (FB, WZ-I, WZ-II); all three showed significant missing in runs checked by the anonymous colleague.

Brier (personal communication, November 25, 1985) interpreted subjects' psi-missing in his experiment as a psi-based reaction against him as contrasted with Feather, their instructor with whom they presumably had developed a relationship. This interpretation does not apply well to our results because we were guest speakers meeting subjects for the first time. In addition, to the extent that below-chance scores are related to checking runs under not-blind conditions, explanations based on subjects' reactions, which are not likely to vary between blind and not-blind runs, are not supported. Similarly, explanations consistent with OT but not taking observational factors into account (e.g., that the checker, as the psi source, had a "psi-missing personality") are contradicted. Further work will be needed to clarify this pattern and to understand the causes of psi-missing generally.

Low Variance in Blind Data

The unusually low C × P interaction in Series II's blind condition was replicated in Series IV. The probability of obtaining such small variance was $< .01$ (two-tailed) for Series II and $< .001$ (two-tailed) for Series IV (Hays, 1973, p. 447). It is unwise to attach meaning to a "significantly nonsignificant" outcome, but this recurrence of low empirical variance in the blind data should at least be briefly addressed.

First, it should be noted that the variance being discussed here is not the same as run-score or subject variance studied by Rogers, Carpenter, and others (Carpenter, 1977). Run-score or subject variance represents deviations around the theoretically expected value (MCE), whereas the variance being tested in an ANOVA represents the spread of scores around the observed (empirical) mean. To the extent that the empirical mean is close to mean chance expectation, these two types of variance will be similar in magnitude but they are conceptually distinct. (See Weiner, 1982, for further discussion.) Second, it is important to bear in mind that the small F ratios are found in interaction terms. Thus, the question to be addressed is not why the scores obtained under a particular condition are similar to each other but, rather, why the pattern of scores among the four blind conditions is similar (see Fig. 1).

As discussed in our previous report, two possible explanations for this low variance are: (1) that an unknown intervening variable systematically reversed the C × P interaction on some runs, cancelling out a C × P interaction in the remaining runs; and (2) that the violation of the laws of probability in the significant not-blind C × P interaction was "balanced" or compensated for by the close fit to these laws in the blind C × P interaction (e.g., Bierman, 1985; Carpenter, 1983; Eisenbud, 1963; Rao, 1978). No intervening variable clearly presents itself as a possible cause for a systematic reversal, though this is not to say that none exists. As to the second possible explanation, it should be noted that in a literature review George (1981) tested whether increased variance in an ANOVA caused by a presumed psi effect was balanced by decreased variance elsewhere (either in the same study or across parapsychological studies) and found no supporting evidence.

The small variance in the present research may be related to conditions surrounding the "second step" of observation, decoding blind runs into prediction categories: significantly low variance was found in the blind conditions decoded by an assistant (Series II, B2 condition of Series IV) but not in the blind condition decoded by the experimenters (Series III, B1 condition of Series IV). A review of second-order ("future") observer effects has shown a trend toward small empirical variances in tests of future-observer effects when data had been constrained by previous observation (Weiner, 1982). Perhaps the assistants' decoding led to some (though not complete) constraint of the data, forcing subsequent observers to influence the data at higher moments of the distribution. This interpretation is, of course, highly speculative, and a mechanism for such influence is not clear. Still, the similarity of our results to trends in the second-order observer literature is of theoretical interest.

Meaningful vs. Complete Observation

In our earlier report we discussed two characteristics of our manipulation of observational context that may have been important: that the checkers' knowledge of the run's prediction category made their observation of call-to-target matches meaningful and provided for complete reduction of uncertainty in one step. What do Series III and IV say about the relative importance of these two factors? Series IV, by replicating Series II, supports the importance of both, for checker effects again occurred only under not-blind conditions, when observation was both meaningful and complete. Series III, however, suggests that the latter plays a more fundamental role, for significant results were found in the "nonmeaningful" blind condition. On the other hand, the blind runs in this series had been decoded by persons for whom the distinction between yes and no had personal significance and who understood how trends in these data would affect the hypotheses under test. Thus, observation during the second step was as meaningful (in the sense used here) as observation of trial-by-trial data had been. Therefore, at present we have no reason for believing that either of these two aspects of observation, meaningfulness or completeness, is necessarily more important than the other.

Conclusions

The present research suggests that critical attention be paid to the act of checking precognition data. The results support the observational model of ESP, which posits that observing the call-to-target match is a crucial factor in ESP test results. Further, they are consistent with

O'Brien's findings regarding checker expectation: checker influence was obtained only under the conditions that allowed checkers to form specific expectations, intentions, and desires for the outcomes. Additional anecdotal support for the efficacy of checker attention and intention comes from the trend toward significant differentiation in N.L.Z.'s not-blind runs as she paid increasingly more attention to run condition and altered her observational strategy to be more consistent with that used successfully by D.H.W. in Series I and II until (in Series IV) she "created" scoring differences in a prespecified direction.

In addition to its primary finding regarding the role of the observational context of data checking, this research shows two interesting trends: (1) psi-missing in data checked by the anonymous colleague, especially under not-blind conditions; and (2) significantly low variance when data undergo a two-step observational process in which the second step is conducted by relatively disinterested parties. Both trends deserve further attention.

Certainly, not all experimental precognition results—and few, if any, precognition experiences—can be attributed to checker effects. Still, viewing precognition studies from alternative viewpoints, such as that of the checker, may help to resolve some of the contradictions and riddles in the literature. Whatever the ultimate outcome of this line of research, it will have served a useful purpose if it stimulates novel and creative ways of conceptualizing psi experiments.

References

Bierman, D. J. (1985). Parapsychology and physics. *Journal of Indian Psychology*, **4** (2), 27–42.

Bierman, D. J., Houtkooper, J. M., & Millar, B. (1981, April). *Triple observation of the observational theories.* Paper presented at the Fifth International Conference of the Society for Psychical Research, Bristol, England.

Braude, S. (1979). The observational theories in parapsychology: A critique. *Journal of the American Society for Psychical Research*, **73**, 329–366.

Braude, S. (1988). Death by observation: A reply to Millar. *Journal of the American Society for Psychical Research*, **82**, 273–280.

Broughton, R. S. (1977). An exploratory study on psi-based subject and experimenter expectancy effects. In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 173–177). Metuchen, NJ: Scarecrow Press.

Carpenter, J. C. (1977). Intrasubject and subject-agent effects in ESP experiments. In B. B. Wolman (Ed.), *Handbook of parapsychology* (pp. 202–272). New York: Van Nostrand Reinhold.

Carpenter, J. C. (1983). Prediction of forced-choice ESP performance: Part I. A mood-adjective scale for predicting the variance of ESP run scores. *Journal of Parapsychology*, **47**, 191–216.

Eisenbud, J. (1963). Psi and the nature of things. *International Journal of Parapsychology,* **5,** 245–273.

Feather, S. R., & Brier, R. (1968). The possible effect of the checker in precognition tests. *Journal of Parapsychology,* **32,** 167–175.

Fisk, G. W., & West, D. J. (1958). Dice-casting experiments with a single subject. *Journal of the Society for Psychical Research,* **39,** 277–287.

George, L. (1981). *Psi as a universal homeostatic function: Toward a testable theory.* Unpublished manuscript.

Hays, W. L. (1973). *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart, & Winston.

Houtkooper, J. M., & Haraldsson, E. (1985). Experimenter effects in a plethysmographic ESP experiment. *European Journal of Parapsychology,* **5,** 313–326.

Kennedy, J. E., O'Brien, J. T., O'Brien, D., & Kanthamani, H. (1977). *An attempt to replicate the checker effect.* Unpublished manuscript. Abstracted in *Journal of Parapsychology,* **41,** 377–378.

Kennedy, J. E., & Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *Journal of Parapsychology,* **40,** 1–33.

Kornwachs, K., & von Lucadou, W. (1979). Psychokinesis and the concept of complexity. *Psychoenergetic Systems,* **3,** 327–342.

Kreiman, N., Ivnisky, D., & Marquez, L. E. (1987). El efecto del revisor ("checker effect"). *Cuadernos de Parapsicología* **20** (1), 1–9.

May, E. C., Radin, D. I., Hubbard, G. S., Humphrey, B. S., Utts, J. M. (1985). Psi experiments with random number generators: An informational model. In *Proceedings of the 28th Annual Convention of the Parapsychological Association,* Vol. 1 (pp. 235–266). Abstracted in D. H. Weiner & D. I. Radin (Eds.), *Research in parapsychology 1985* (pp. 119–120). Metuchen, NJ: Scarecrow Press, 1986.

Millar, B. (1988). Cutting the Braudian loop: In defense of the observational theories. *Journal of the American Society for Psychical Research,* **82,** 253–271.

Morris, R. L. (1968). Obtaining non-random entry points: A complex psi task. In J. B. Rhine & R. Brier (Eds.), *Parapsychology today* (pp. 75–86). New York: Citadel.

O'Brien, J. T. (1979). An examination of the checker effect. In W. G. Roll (Ed.), *Research in parapsychology 1978* (pp. 153–155). Metuchen, NJ: Scarecrow Press.

Phillips, P. (1984). Measurement in quantum mechanics. *Journal of the Society for Psychical Research,* **52,** 297–306.

The RAND Corporation (1955). *A million random digits with 100,000 normal deviates.* Glencoe, IL: Free Press.

Rao, K. R. (1978). Psi: Its place in nature. *Journal of Parapsychology,* **42,** 276–303.

Rhine, J. B., & Pratt, J. G. (1957). *Parapsychology: Frontier science of the mind.* Springfield, IL: Charles C Thomas.

Schmidt, H. (1975). Toward a mathematical theory of psi. *Journal of the American Society for Psychical Research,* **69,** 301–319.

STANFORD, R. G. (1981). Are we shamans or are we scientists? *Journal of the American Society for Psychical Research,* **75,** 61–70.
VON LUCADOU, W., & KORNWACHS, K. (1977a). Can quantum theory explain paranormal phenomena? In J. D. Morris, W. G. Roll, & R. L. Morris (Eds.), *Research in parapsychology 1976* (pp. 187–191). Metuchen, NJ: Scarecrow Press.
VON LUCADOU, W., & KORNWACHS, K. (1977b, September). *Fitting of PK results with a quantum mechanical model?* Paper presented at the 1977 Parascience Conference, London, England.
VON LUCADOU, W., & KORNWACHS, K. (1980a). Development of the system theoretic approach to psychokinesis. *European Journal of Parapsychology,* **3,** 297–314.
VON LUCADOU, W., & KORNWACHS, K. (1980b, September). *On the structure of observational theories.* Paper presented at the 1980 Parascience Conference, London, England.
VON LUCADOU, W., & KORNWACHS, K. (1983). On the limitations of psi: A system-theoretic approach. *Psychoenergetics,* **5,** 53–72.
WALKER, E. H. (1975). Foundations of paraphysical and parapsychological phenomena. In L. Oteri (Ed.), *Quantum physics and parapsychology* (pp. 1–44). New York: Parapsychology Foundation.
WALKER, E. H. (1984). A review of criticisms of the quantum mechanical theory of psi phenomena. *Journal of Parapsychology,* **48,** 277–332.
WEINER, D. H. (1982, February). *The once and future psi source: A review of the literature on future-observer effects.* Paper presented at the Southeastern Regional Parapsychological Association Conference, Chapel Hill, NC. Abstracted in *Journal of Parapsychology,* **46,** 58–59.
WEINER, D. H., & ZINGRONE, N. L. (1986). The checker effect revisited. *Journal of Parapsychology,* **50,** 85–121.
WEST, D. J., & FISK, G. W. (1953). A dual ESP experiment with clock cards. *Journal of the Society for Psychical Research,* **37,** 184–197.

Institute for Parapsychology
Box 6847 College Station
Durham, NC 27708-6847