Psychological Bulletin
1994, Vol. 115, No. 1, 19–24

0033-2909/94/$3.00

Anomaly or Artifact? Comments on Bem and Honorton

Ray Hyman

Bem and Honorton imply that the 11 autoganzfeld experiments demonstrate the existence of psi—a communications anomaly. They claim that the autoganzfeld results are consistent with previous parapsychological findings and constitute evidence for a replicable psi effect. Although the autoganzfeld experiments are methodologically superior to previous parapsychological experiments, the tests of their randomization procedures were inadequate. The autoganzfeld experiments consistently produced positive hit rates, whose combined effect was highly significant. However, these experiments produced important inconsistencies with the previous ganzfeld experiments. They also showed a unique pattern in the data that may reflect a systematic artifact. Because of these unique features, we have to wait for independent replications of these experiments before we can conclude that a replicable anomaly or psi has been demonstrated.

Bem and Honorton (1994) imply that if psychologists were familiar with the most recent parapsychological research, they would be more willing to accept the possibility that a communications anomaly existed. In particular, Bem and Honorton focus on the experiments that are based on the ganzfeld procedure. They "believe that the replication rates and effect sizes achieved by one particular experimental method, the ganzfeld procedure, are now sufficient to warrant bringing this body of data to the attention of the wider psychological community" (Bem & Honorton, 1994, p. 4). They review the debate between Honorton and me over the original ganzfeld experiments. Hyman (1985) found that these studies suffered from statistical, methodological, and documentation problems. Honorton (1985) responded that these flaws were not sufficient to account for the observed hit rates. Bem and Honorton (1994) review this controversy and cite reviewers who apparently agree with Honorton's position. The implication is that despite the deficiencies in the ganzfeld experiments, the results support the existence of psi—a communications anomaly.

To Honorton's credit, he initiated a new series of experiments that would be free from the flaws of the earlier ganzfeld database (Honorton et al., 1990). These 11 new experiments, called the *autoganzfeld studies* yielded consistently positive hit rates and a highly significant overall effect. Because these new experiments showed positive results and allegedly were consistent with the earlier ganzfeld database and other psi research, Bem and Honorton implied that parapsychology had found its previously elusive repeatable experiment.

Since the beginnings of psychical research in the mid-nineteenth century, its investigators have believed that they have scientific evidence sufficiently strong to place before the general scientific community. Each generation has tried to get the attention of the scientific community with findings that they claim to be irrefutable. The particular evidence put forth has changed from generation to generation. What a previous generation of parapsychologists considered to be a solid case for psi was abandoned by later generations in favor of a more current candidate. This shifting database for parapsychology's best case may be why parapsychology still has not achieved the recognition it desires from the general scientific community.

Now Bem and Honorton (1994) believe that they have a strong case to put before the psychological community. They admit that the autoganzfeld findings still require independent confirmation. To their credit, they specify the conditions and the required sample size needed to provide adequate power. The informed critic of parapsychology might ask what makes the current situation different from the past claims for psi? Why should we now believe that Honorton and his colleagues have finally found a way to consistently produce evidence for psi?

We must wait for future attempts at replication before we have an answer to the question. Bem and Honorton appear confident that this time is different. Their review of the ganzfeld and autoganzfeld databases encourages them to believe that consistent psi results are within reach. In this commentary, I provide reasons for believing that the autoganzfeld results contain inconsistencies and some unique patterns that raise doubts about their replicability.[1]

Agreements and Differences

Although my commentary focuses on my disagreements with Bem and Honorton's (1994) presentation, I would like to briefly specify some points of agreement. The autoganzfeld studies do comply with most of the "stringent standards" (p. 353) spelled out in the joint communiqué by Hyman and Honorton (1986). I commend Honorton and his colleagues (1990) for creating a protocol that eliminates most of the flaws that plagued the original ganzfeld experiments. The 11 autoganzfeld studies consistently yield positive effects that, taken together, are highly significant. I concur with Bem and Honorton's admission that

Correspondence concerning this article should be addressed to Ray Hyman, Department of Psychology, University of Oregon, Eugene, Oregon 97403.

[1] Although I take a pessimistic position regarding future replications, I think it is good that Bem and the parapsychologists are optimistic. Such optimism should encourage investigators to attempt replications. These replications will eventually decide the issue.

"the autoganzfeld studies by themselves cannot satisfy the requirement that replications be conducted by a 'broader range of investigators' " (p. 13). I also support their suggestion that several parapsychologists pool their resources and plan a large-scale ganzfeld replication in which each laboratory contributes a set of trials to the total pool.

So what is there to disagree about? I disagree with Bem and Honorton about how strongly the autoganzfeld studies support the hope for a replicable psi experiment. Where they see consistency between the autoganzfeld studies and previous parapsychological findings, I see inconsistency. Although I agree that the autoganzfeld studies meet most of the stringent standards that Honorton and I spelled out, I disagree that they meet all of those standards. Our disagreements are a matter of degree. The value of discussing our disagreements is to help clarify what should constitute adequate evidence for the existence of an anomaly. The existence or nonexistence of psi will not be settled by debate. The existence issue will be settled by independent attempts at replication—at least four of which are currently underway (McCrone, 1993).

In explaining my disagreements, I point to weaknesses in the autoganzfeld experiments. I want to emphasize that as a single contribution to the ganzfeld database, these are commendable experiments of high quality. But no single experiment or set of studies can be perfect in all respects. When such a series is given the responsibility of carrying a burden beyond its original purposes, then various deficiencies will inevitably become apparent. This is the case, I believe, with the autoganzfeld studies.

Internal Consistency Within the Autoganzfeld Studies

Bem and Honorton describe the autoganzfeld studies as 11 separate experiments conducted by eight different experimenters. The hit rates are positive and consistent across the studies and the experimenters. Although this is encouraging, the consistency tells us little about potential replicability. Neither the studies nor the experimenters are independent. The studies vary in whether they use naive or experienced subjects. However, the target set, the selection and judging procedures, the laboratory, the setting, and the procedures are identical across studies and experimenters. No experimenter is associated with a single study, nor does an experimenter have independent input into the design and procedure as happens in an independent replication. Indeed, the term *experimenter* in this context simply refers to a person who plays an already scripted role. Any unique features of the autoganzfeld procedure—including possible artifacts—would be the same for all 11 studies and the eight different experimenters. Consequently the autoganzfeld studies should be looked on as 1 large experiment rather than 11 separate contributions.

Consistency With the Original Ganzfeld Database

Bem and Honorton claim that "[the autoganzfeld] results are statistically significant and consistent with those in the earlier database" (p. 13). They cite only two reasons to support this claim. The overall effect size or hit rate is approximately the same in the two databases. This apparent agreement in overall effect size is meaningless. The overall effect size in the autoganzfeld studies is a composite of two significantly different effect sizes—that for the static targets and that for the dynamic targets. The overall effect size in the ganzfeld data base is an arbitrary composite of heterogeneous effect sizes, contributed in unequal numbers, from different laboratories. The fact that the two composites yield approximately the same effect size is accidental. Both numbers could easily have been larger or smaller, depending on the mix of the arbitrary sources from which they were composed.

The dynamic targets yielded a significantly higher hit rate than did the static targets in the autoganzfeld studies. Bem and Honorton argued that this was consistent with the finding that the multiple-image targets (View Master stereoscopic slide reels) in the ganzfeld database yielded significantly higher hit rates than did the single-image targets. I do not believe that multiple static images on a View Master reel can be equated to the dynamic moving image on a videoclip. However, I will not argue this point.

Clearly the dynamic targets outperform the static targets in the autoganzfeld studies. Even if this is consistent with the apparent superiority of the View Master targets over the single-image targets, Bem and Honorton (1994) overlook a serious discrepancy. Single-image targets constituted 76% of the 835 sessions in the ganzfeld experiments. Their average hit rate was .346. Given this effect size and the 166 trials using static targets, the power or probability of replicating this effect in the autoganzfeld experiments was .82. This failure to find a significant effect with the static targets was even more notable given that these experiments were conducted in "the warm social ambience" (p. 14) of Honorton's laboratory.

Bem and Honorton acknowledge that the autoganzfeld studies failed to replicate the predicted sender–receiver pairing effect. In the original ganzfeld database, the trials on which the receiver chose a friend as a sender produced a hit rate of .44 compared with a hit rate of only .26 for those trials on which the experimenter assigned a sender. I would emphasize that given this size effect with the 198 trials with friends as senders and 128 with someone else as senders, the power of getting a significant replication of the effect is over .92. Again, given the 'psi conducive' atmosphere of Honorton's laboratory, this failure to get significance is a noteworthy inconsistency.

On two key comparisons with the original ganzfeld database, the autoganzfeld fails to replicate even with adequate power. The positive hit rate and overall significance of the autoganzfeld studies are due to an essentially new type of target, presented in a new way. Even if we agree that there is a kinship between the View Master reels of the ganzfeld experiments and the dynamic targets of the autoganzfeld, we cannot ignore the differences between multiple images of a travel scene presented statically with a slide projector and excerpts from motion pictures presented with their accompanying audio on videocassettes. The problems of selecting, presenting, and controlling such targets present new challenges. During the judging procedure in the original ganzfeld experiments, the target and the decoys were displayed simultaneously. The judging procedure for the autoganzfeld involves presenting the target and its decoys one at a time. Because the positive hit rate and significance are due to an essentially new type of target presented in a new way, the need for independent replication is especially urgent.

Consistency With Previous Parapsychological Findings

Bem and Honorton (1994) also claimed that "there are reliable relationships between successful psi performance and conceptually relevant experimental and subject variables, relationships that also replicate previous findings" (p. 13). They point to three such "replications." One is a small, but statistically significant, correlation of .18 between a measure of extroversion and "psi performance." This is consistent with a tendency found in previous psi studies. Second, they report the strong psi performance of the Julliard students that they see as consistent with psi studies that found a relation between psi abilities and creative and artistic abilities. This latter replication is not so impressive when one considers that only 20 students were involved and that their performance was not significantly different from the other participants in the two studies in which they participated (Fisher's exact $p = .262$, two-tailed). In addition, as I point out below, the Julliard students were exposed to just those conditions that favored high hit rates—targets that were repeated, a preponderance of dynamic targets, and active prompting by the experimenter during judging. Thus, it is unclear whether their high hit rate was a function of their creativity or a function of the special targets and conditions with which they happened to be associated.

The third correlate could not be demonstrated for the autoganzfeld studies. Bem and Honorton (1994) pointed out that the subjects in the autoganzfeld tended to believe in psi, reported psychic experiences, and had practiced meditation or related techniques. These variables were previously reported as correlates of psi. However, I do not see how they can claim that these attributes of their subject population are a replication of previous findings. They report no correlations between these variables and performance in the autoganzfeld studies. Indeed, they cannot report any correlation because they did not have subjects who lacked these properties. We do not know if nonbelievers and people without psychic experiences would have performed better or worse than the actual subjects.

In other words, they can justify only one of the correlates that they use to claim consistency with previous psi studies. Even here the relationship is weak and is just one of many previously reported correlates that might have been found. At one time, for example, parapsychologists claimed that the decline effect was a pervasive and characteristic property of psi. However, when no decline effect is found in a parapsychological study, it does not deter the experimenter from pointing to some other significant departure from chance as evidence for psi. Note that in the autoganzfeld studies, there is no decline effect.

Randomization and Claims of Psi

As I already stated, I agree that the autoganzfeld studies meet most of the requirements that Honorton and I specified in our joint communiqué (Hyman & Honorton, 1986). One surprising exception is the inadequate testing of the randomization procedures. The issue of randomization was central to the debate concerning the original ganzfeld findings (Hyman, 1985). Adequate randomization procedures are critical for parapsychological research because the evidence for psi is based on a low probability value for a departure from a chance baseline. Such probability values have meaning within an idealized statistical model of the experimental situation. Whether this statistical model applies to a given situation is an empirical matter that must be adequately justified if the stated significance levels are to be taken seriously. Appropriate randomization procedures are one way to help ensure that the statistical model applies to the experimental data. With respect to the autoganzfeld studies, this would entail selecting the targets for each trial and ordering the target and decoys during judging in a demonstrably random manner. In addition, following the practice of a few past researchers, the parapsychologist can also provide some post hoc analyses to show that the distributions of targets and judging orders are consistent with the underlying probability model.

Unfortunately, the autoganzfeld studies fell short on this critical requirement. The tests for adequacy of randomization were confined to showing a uniform distribution of outputs from 1 to 160 for target selection and a uniform distribution of the permutations of all possible orderings during the judging procedure. Emitting a uniform distribution of target choices is a necessary but hardly sufficient requirement for an adequate random generator.

These randomization procedures are critical because we can expect strong systematic biases during the judging procedure. The fact that the items to be judged have to be presented sequentially, when combined with what we know about subjective validation (Marks & Kammann, 1980), would lead us to expect a strong tendency to select the first or second items during the judging series. We would also expect strong response biases within each target pool. Bem and Honorton show such a bias in the target pool used for Study 302. Both these biases may be strengthened by the fact that the experimenter interacts with the receiver during the judging process. Although most receivers participate in one session, each experimenter participates in several. The response biases of the experimenters can play an important role, especially in those studies in which the experimenter deliberately prompts the receiver to choose a particular item during the judging. Such active prompting occurred in 6 of the 11 studies (Honorton et al., 1990).[2]

If the randomizing of the selection of targets and of the ordering of items during judging is adequate, such response biases should not affect the validity of the statistical tests. One way to prevent response biases from distorting the hit rate is to use a randomizing procedure that makes sure that each item within a target pool occurs equally often. The simple randomizing procedure used in the autoganzfeld studies would guarantee that each target occurred an equal proportion (not number) of times only in the very long run. In any finite number of trials, the individual targets would occur with varying frequencies. Again, if the randomization was adequate, this inequality of occurrence would not bias the hit rate. The items in some target pools that occurred most frequently would be those that were favored

[2] One referee suggested that I make it clear that I am not claiming that sensory leakage occurred because of experimenter prompting. I agree. The experimenter, according to the protocol, was ignorant of which member of the target pool was the target during the judging procedure. The point is that by actively helping the subject to rate the members of the target pool, the experimenter let his or her own subjective biases enter the selection procedure.

Table 1
Hit Rate as a Function of the Frequency of Occurrence of Targets

	Frequency								
Variable	1	2	3	4	5	6	7	8	Total
Hits	12	25	16	20	19	4	4	6	106
Misses	36	65	26	48	36	8	3	2	224
n	48	90	42	68	55	12	7	8	330
Hit rate	.250	.278	.381	.294	.346	.333	.571	.750	.321

by the response bias. This would bias the hit rate upward. The items in other target pools, however, that occurred most frequently would be those that were avoided by the response bias. This would bias the hit rate downward. With adequate randomization, these two tendencies would balance each other.

Achieving adequate randomization is not easy. Much can go wrong—as some parapsychologists, among others, have shown. This is why it is disappointing that the autoganzfeld studies did not show the same concern for randomizing that they showed for other aspects of the methodology. This is also why, in my role of devil's advocate, I was interested in directly checking the actual distribution of target positions among the decoys during judging. Daryl Bem kindly agreed to supply me with this information along with other data from the autoganzfeld database. Unfortunately, the variable labeled *position* on the data sheet turned out to be the original position of the target in its target pool rather than its position during judging. This latter information was unavailable to either Bem or me at the time of this writing.

Hit Rate and Target Frequency

Because I could not directly check the adequacy of the randomization procedures, I tried to find some indirect indicators. If randomizing was inadequate and targets occurred with varying frequency, possible biases might show up as differential hit rates for targets occurring with various frequencies. For example, if targets favored by response biases were also favored by a deficient target selection procedure, then we would find a positive correlation between hit rate and target frequency. It would be possible, of course, for a deficient randomization procedure to yield a negative correlation. To see if actual repetitions of targets had any observable consequences, I tabulated the proportion of hits as a function of how many times a target occurred in this database.[3]

As Table 1 shows, the relation between hit rate and target frequency was strong. The test for a linear trend among the proportions (Snedecor & Cochran, 1967, pp. 246–248) was positive and significant, ($z = 2.49$, $p = .013$, two-tailed). An indication of the strength of this trend is given by the Spearman rank order correlation between the hit rate and target frequency, which was .83. Another way to look at this relationship would be to compare the hit rate of targets that occurred once or twice (.27) with those that appeared three or more times (.36).

This pattern exists separately for the static and dynamic targets, although it is stronger among the dynamic targets. The static targets that occurred once or twice had a hit rate of .22 compared with a hit rate of .31 for those that occurred more than twice. The hit rate was .32 for those dynamic targets that occurred once or twice as compared with a hit rate of .41 for those that occurred three or more times.

Target Occurrence and Experimenter Prompting

What accounts for this peculiar relationship? Is the correlation between target frequency and hit rate determined by which particular targets get repeated? Or does replication itself somehow increase the hit rate? If the relation is due to response biases, I would expect experimenter prompting to affect the later occurrences of targets rather than their first occurrences. With these questions in mind, I conducted a multinomial analysis of variance (Woodward, Bonett, & Brecht, 1990). In this analysis, hit rate was the dependent variable, and 3 two-level factors were the independent variables: target type (static, dynamic), target occurrence (first, later), and experimenter prompting (no, yes). Of the interactions, only that between target occurrence and experimenter prompting was significant, $\chi^2(1, N = 330) = 6.83$, $p = .009$. The two significant main effects were target type, $\chi^2(1, N = 330) = 4.76$, $p = .030$, and target occurrence, $\chi^2(1, N = 330) = 11.56$, $p < .001$.

The difference between the hit rate for dynamic targets (.356) and that for static targets (.249) does not interact with the other two factors and will be ignored for the present discussion.[4] The meaning of the interaction between target occurrence and prompting can be seen in the simple effects of target occurrence within each level of prompting. With no experimenter prompting, the effects of target occurrence were minimal: The hit rate for first occurrences of targets was .291 and that for later occur-

---

[3] To be consistent with Bem and Honorton, I treated the basic database as the 330 sessions in Studies 1 through 301. Study 302, which used a single target pool, was treated as a special case.

[4] These hit rates are slightly different from those used by Honorton et al. (1990) and Bem and Honorton (1994). This is because they computed hit rates for any category by simply dividing the number of hits by the total number in that category. The hit rate for dynamic targets obtained with this approach is 61/164 = .372 and that for static targets is 45/166 = .271. These rates are means weighted by the number of cases in the cells for each combination of levels of the factors. For the purposes of additivity of effects, I am using the unweighted means (each cell of the design is weighted equally). This removes distortions and confounds that are due to unequal cell sizes. In the present case, the differences are small and inconsequential. I am supplying this footnote to explain some discrepancies that might confuse the reader.

rences was .334, $\chi^2(1, N = 181) = .396, p = .534$. The effect of target repetition combined with experimenter prompting, however, was very large. The hit rate for first occurrences of targets with experimenter prompting was only .140. The hit rate for later occurrences of targets when combined with experimenter prompting jumped to .445. This gain was significant, $\chi^2(1, N = 149) = 14.702, p = .0001$. These results suggest that experimenter prompting depresses hit rates for first occurrences of targets and enhances hit rates for subsequent occurrences of targets.

Internal Checks on the Validity of This Pattern

Is this peculiar relation among hit rate, target frequency, and experimenter prompting merely a fluke? I broke the data into subsets in several ways to see if the pattern was consistent in the different subcategories. I checked this pattern within the dynamic and static targets separately. I compared Targets 1 to 80 with Targets 81 to 160. Likewise, I looked for the pattern within Studies 101 through 103 taken as a group as compared with Studies 104 through 301 considered as a group. I also checked for this pattern for each of the five experimenters who contributed the most sessions. Although the numbers became small in some of these comparisons, the hit rate was consistently larger for later as opposed to first occurrences of a target. I found just one nonsignificant exception in the trials for one experimenter. Likewise, wherever meaningful comparisons were possible, the interaction between prompting and target occurrence occurred.[5] For this database, then, the dependence of hit rate on frequency of target occurrence and experimenter prompting was a robust effect.

Implications

As far as I know, this dependence of hit rate on target occurrence and experimenter coaching has never been previously reported in parapsychological research. One referee suggests that the dependence of hit rate on target frequency and prompting may reveal important moderator variables rather than artifacts. The referee may be correct. The skeptic, however, might point to the long history of alleged "moderator" variables in parapsychology—such as the decline effect, displacement effects, sheep–goat effects, and others. The problem is that when such moderators are discovered in the data, they are put forth as important indicators or characteristics of psi. The absence of such characteristics in subsequent data, however, does not deter parapsychologists from claiming evidence for psi if they find a significant hit rate. This is the troublesome problem of boundary conditions. The parapsychologists have been unable to specify what would constitute the absence of psi.

The positive effect for repeated occurrences of a target may eventually turn out to be an important property of psi—if psi exists. However, the fact that first occurrences of a target produce a hit rate consistent with chance raises questions. All of the positive effect in the ganzfeld experiments rests on those targets that have occurred more than once. The prompting effect is even more curious. On first occurrences of a target, active coaching by the experimenter seems to depress the hit rate—.28 without prompting versus .15 with prompting. For second or later occurrences of a target, active coaching appears to enhance the hit rate—.33 without prompting versus .45 with prompting. If the prompting by the experimenter is intended to increase reliability by reminding the receiver of ganzfeld associations that he or she might overlook during judging, why should the effects of such prompting show up only for the subsequent occurrences of a target?

That hit rate correlates with frequency of target occurrence could mean that the "better" targets are somehow selected more often by the randomizing procedure. Or it could mean that frequency of occurrence, itself, is the determinant of a higher hit rate. The data suggest the latter possibility. The 48 targets that occurred exactly once in the database had a hit rate of .22. The first occurrence of the targets that occurred more than once had a hit rate of .23. The combined hit rate for second or later occurrences of targets was .39. Another way of examining this relation would be to look separately at the hit rates for first and subsequent occurrences of targets that appeared exactly twice, three times, four times, and so on. Only the targets that occurred from two to five times could be used because only one or two targets appeared with frequencies of six or more. In all these comparisons, the first occurrences consistently had a lower hit rate than subsequent occurrences of the same targets.

Whatever the source for this pattern, it raises questions about interpretations of other findings in the database. For example, Bem and Honorton (1994) pointed to the high rate of .50 for the 20 Julliard students as evidence for the effect of artistic creativity on hit rate. However, all of the sessions in which the Julliard students appeared were prompted, and 15 of the 20 used second or later occurrences of a target. On the five targets that were occurring for the first time, these students got one hit. Consequently, we cannot tell if the hit rate for these students reflect any special abilities or if they are due to whatever makes hit rate a function of target frequency and coaching in this database.

Are these findings due to an artifact, or do they point to some new, hitherto unrecognized property of psi? We cannot say. The existence of this pattern in the database, however, strongly supports the need to replicate the findings before we can be confident that the parapsychologists have finally found a way to capture and tame their elusive quarry.

Conclusions

The autoganzfeld experiments are a praiseworthy improvement in methodological sophistication and experimental rigor over the previous ganzfeld experiments. Despite these improvements, the experiments fall disappointingly short in the critical area of justifying the randomization procedures. Even though all but one of the individual studies produced a positive effect size and the overall effect was significant, the autoganzfeld experiments do not constitute a successful replication of the original ganzfeld experiments.

---

[5] As noted in Footnote 5 of the Bem-Honorton article, a recent review of the original computer files uncovered a duplicate record in the autoganzfeld database. This has now been eliminated, reducing by one the number of sessions on which my analysis was based. Some experimenters contributed only unprompted sessions, and some contributed mainly prompted sessions.

Although Bem and Honorton point to consistencies between the autoganzfeld results and those of previous parapsychological research, these consistencies are more apparent than real. On the other hand, as I have argued, important inconsistencies exist between the two databases.

Three robust effects in the autoganzfeld database are the dependence of hit rate on type of target (dynamic or static), target occurrence (first or subsequent), and experimenter prompting (yes or no). Although my looking for effects of the latter two factors was motivated by my concern for possible randomization deficiencies, their existence should interest both parapsychologists and critics. This is because the existence of an effect depends on these factors. The combination of dynamic targets, repeated occurrences of a target, and experimenter prompting produces a hit rate of .471 with 95% confidence limits from .305 to .629. The combination of static targets, first occurrences of a target, and no prompting yields a hit rate of .178 with 95% confidence limits from .066 to .336.

We do not have enough information to know if the dependence of hit rate on target frequency and experimenter prompting involves response preferences for items within a target pool. One way to ensure that such preferences do not bias the hit rate is to present each member of a target pool equally often. I tried to get some idea what the hit rate might be if each member of a target pool had occurred equally often. I restricted myself just to those target pools in which each member occurred at least once.[6] The hit rate for first occurrences of a target in these target pools was .275 with 95% confidence limits from .167 to .399. (The hit rate for the targets in these target pools that occurred a second or later time was .427.) This finding does not prove anything, but it suggests that if the targets within each target pool had occurred equally often, the results might have been consistent with chance.

The autoganzfeld studies failed to replicate key findings of the original ganzfeld experiments, even though the power was sufficient. The positive effect size and significance depended on a new type of target whose presentation involves a new technology and on target repetition and experimenter coaching. Whatever their source, these effects are new to the psi literature. We do not know how much of this is unique to this experimental setup and laboratory. For these reasons, we have to wait for future attempts at replication to see if a replicable psi effect is at hand.

---

[6] I could not use higher frequency of occurrence because only three target pools existed in this database that had at least two occurrences of each of its members.

References

Bem, D.J., & Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin, 115,* 4–18.

Honorton, C. (1985). Meta-analysis of psi ganzfeld research: A response to Hyman. *Journal of Parapsychology, 49,* 51–91.

Honorton, C., Berger, R.E., Varvoglis, M.P., M., Derr, P., Schechter, E.I., & Ferrari, D.C. (1990). Psi communication in the ganzfeld: Experiments with an automated testing system and a comparison with a meta-analysis of earlier studies. *Journal of Parapsychology, 54,* 99–139.

Hyman, R. (1985). The ganzfeld psi experiment: A critical appraisal. *Journal of Parapsychology, 49,* 3–49.

Hyman, R., & Honorton, C. (1986). A joint communiqué: The psi ganzfeld controversy. *Journal of Parapsychology, 50,* 351–364.

Marks, D., & Kammann, R. (1980). *The psychology of the psychic.* Buffalo, NY: Prometheus Books.

McCrone, J. (1993, May 1). Roll up for the telepathy test. *New Scientist,* pp. 29–33.

Snedecor, G.W., & Cochran, W.G. (1967). *Statistical methods* (6th ed., pp. 246–248). Ames: Iowa State University Press.

Woodward, J.A., Bonett, D.G., & Brecht, M. L. (1990). *Introduction to linear models and experimental design.* San Diego, CA: Harcourt Brace Jovanovich.

Received July 26, 1993
Accepted July 27, 1993 ■