PSI PHENOMENA

Daryl J. Bem

Cornell University

Prepared for inclusion in Atkinson, R, L, Atkinson, R. C., Smith, E. E., & Bem, D. J. (1990). *Introduction to Psychology* (10th Ed). San Diego: Harcourt Brace Jovanovich. I am grateful to Charles Honorton, Robert Rosenthal, and Marcello Truzzi for comments on earlier drafts.

Approved For Release 2000/08/10 : CIA-RDP96-00792R000100130003-0

PSI PHENOMENA

Is it possible for us to acquire information about the world in ways that do not involve stimulation of the known sense organs or to influence physical events by purely mental means? These questions are the source of controversy over the existence of *psi*, processes of information and/or energy exchange not currently explicable in terms of known science. The phenomena of psi are the subject matter of parapsychology ("beside psychology") and include the following:

1. **Extrasensory Perception (ESP).** Response to external stimuli without sensory contact.
   a. **Telepathy.** Response to the mental state of another person without the mediation of any known channel of sensory communication (for example, identifying a playing card merely being thought of by another person)
   b. **Clairvoyance.** Response to objects or events that do not provide a stimulus to any known sense (for example, identifying a concealed playing card whose identity is unknown to anyone)
   c. **Precognition.** Response to a future event that could not be anticipated through any known inferential process (for example, predicting which digit a random number generator will generate on the next trial)
2. **Psychokinesis (PK).** Mental influence over physical events without the mediation of any known physical force (for example, influencing which digit a random number generator will generate on the next trial)

Experimental Evidence

Most parapsychologists consider themselves to be scientists applying the usual rules of scientific inquiry to admittedly unusual phenomena. Yet the phenomena of psi are so extraordinary and so similar to what are widely regarded as superstitions that some scientists

declare psi to be an impossibility and reject the legitimacy of parapsychological inquiry. Such a priori judgments are out of place in science; the real question is whether the empirical evidence is acceptable by scientific standards. Many psychologists who are not yet convinced that psi has been demonstrated are nevertheless open to the possibility that new evidence might emerge that would be more compelling. For their part, many parapsychologists believe that several recent experimental procedures either provide that evidence already or hold the potential for doing so. We shall examine one of the most promising of these, the ganzfeld procedure.

The ganzfeld procedure tests for telepathic communication between a subject acting as the "receiver" and another person who serves as the "sender." The subject is sequestered in an acoustically-isolated room and placed into a mild form of perceptual isolation: translucent ping-pong ball halves are taped over the eyes and headphones are placed over the ears; diffuse red light illuminates the room, and white noise is played through the headphones. (White noise is a random mixture of sound frequencies similar to the hiss made by a radio tuned between stations.) This homogeneous visual and auditory environment is called the *Ganzfeld*, a German word meaning "total field."

The sender is sequestered in a separate acoustically-isolated room, and a visual stimulus (picture, slide, or brief videotaped sequence) is randomly selected from a large pool of similar stimuli to serve as the "target" for the session. While the sender concentrates on the target, the subject attempts to describe it by providing a continuous verbal report of his or her ongoing imagery and free associations. Upon completion of the session, the subject is presented with four stimuli—one of which is the target—and asked to rate the degree to which each matches the imagery and associations experienced during the ganzfeld period. A "direct hit" is scored if he or she assigns the highest rating to the target. (In some studies, individuals unconnected with the experiment also attempt to match a transcript of the subject's verbal report to the target.)

More than 50 ganzfeld experiments have been conducted since the procedure was first introduced in 1974. An overall analysis of 28 studies reported through 1981—comprising 835

ganzfeld sessions conducted in 10 different laboratories—reveals that subjects selected the target 38% of the time on the average. (Because there are 4 alternatives, we would expect a success rate of 25% if only chance were operating.) Statistically this result is highly significant: the probability that it could have arisen by chance is less than one in a billion (Honorton, 1985).

Debate over the Evidence

In 1985 and 1986, the *Journal of Parapsychology* published an extended examination of the ganzfeld studies, focusing on a debate between Ray Hyman, a cognitive psychologist and critic of parapsychology, and Charles Honorton, a parapsychologist and major contributor to the ganzfeld database. They agree on the basic quantitative results but disagree on points of interpretation (Hyman, 1985; Hyman & Honorton, 1986; Honorton, 1985). We shall use their debate as a vehicle for examining the major issues involved in evaluating all claims of psi.

Replication Problem In science generally, a phenomenon is not considered established until it has been observed repeatedly by several researchers. Accordingly, the most serious criticism of parapsychology is that it has failed to produce a single demonstration of psi that can be reliably replicated (successfully reproduced) by other investigators. Even the same investigator testing the same individuals over time may obtain significant results on one occasion but not on another. The ganzfeld procedure is no exception; fewer than half (43%) of the 28 studies analyzed in the debate yielded statistically significant results.

The parapsychologists' most effective response to this criticism actually comes from within psychology itself. Many statisticians and psychologists are dissatisfied with psychology's focus on the significance level as the sole measure of a study's success. As an alternative, they are increasingly adopting the technique of *meta-analysis*, a technique that treats the accumulated studies of a particular phenomenon as a single grand experiment and each study as a single observation. Thus any study that obtains results in the positive direction—even though it may not be statistically significant itself—contributes to the overall strength and

reliability of the phenomenon rather than simply being dismissed as a failure to replicate (Glass, McGaw, & Smith, 1981; Rosenthal, 1984).

From this perspective, the ganzfeld studies provide impressive replicability: 23 (82%) of the 28 studies obtain positive results (more direct hits than chance would predict), a result whose probability of occurring by chance is less than 1 in a thousand.

The conventional criterion of replication further requires that any competent investigator be able to reproduce the claimed phenomenon, not just one or two gifted experimenters. This is often a difficult criterion to achieve in new areas of investigation because a number of unsuspected variables might affect the outcome. In psychological experiments, the experimenter is often an important social stimulus for the subject and hence a poorly controlled source of variability. Even in such an established area as classical conditioning, investigators at one university were obtaining positive results 94% of the time while other investigators could do so only 62% of the time (Rosenthal, 1966; Spence, 1964). Nor is psychology alone here. Similar replication difficulties have been reported in medical studies of placebo efficacy (Moerman, 1981) and in the physical science area of laser technology (Collins, 1974).

This problem may be even more acute in parapsychology because psi effects may legitimately depend upon the motivational atmosphere established by the experimenter. Some parapsychologists further believe that the experimenter's own psi abilities and/or attitudes can affect the results.

Despite these potential difficulties, the replicability of the ganzfeld effect does not appear to rest on the success of one or two investigators. Six of the 10 investigative teams contributing to the 28 examined studies obtained statistically significant results; and, even if all the studies from the two most successful laboratories are discarded from the analysis (half of the studies), the results remain significant (Palmer, Honorton, & Utts, 1988).

The power of a particular experiment to replicate an effect also depends upon how strong the effect is and how many observations are made. If an effect is weak, an experiment

with too few subjects, sessions, or observations will fail to detect it at a statistically significant level—even though the effect actually exists.

This is strikingly illustrated by a recent medical experiment designed to determine whether aspirin can prevent heart attacks. The study was discontinued in 1987 because it was already clear the answer was yes. After six years, the aspirin group had already suffered 45% fewer heart attacks than a control group which received only placebo medication, a result that would occur by chance less than one time out of a million (The Steering Committee of the Physicians' Health Study Research Group, 1988). With such impressive results, it was considered unethical to keep the control group on placebo medication. The study was widely publicized as a major medical breakthrough.

The pertinent point here is that the study included over 22,000 subjects. If it were to be repeated with 3,000 subjects, a significant aspirin effect would be unlikely to emerge; the experiment would fail to replicate. Despite its undisputed reality and its practical importance, the aspirin effect is actually quite weak.

Now reconsider the ganzfeld effect. If the effect actually exists and has a true direct-hit rate of 38%, then statistically we should expect ganzfeld studies with 30 sessions (the average for the 28 studies) to obtain a significant psi effect only about one-third of the time (Utts, 1986). The ganzfeld effect is about 4 times stronger than the aspirin effect.

In short, it is unrealistic to demand that any "real" effect be replicable at any time by any competent investigator. The replication issue is more complex than that, and meta-analysis is proving to be a valuable tool for dealing empirically with some of those complexities.

Inadequate Controls The second major criticism of parapsychology is that many, if not most, of the experiments have inadequate controls and safeguards. Flawed procedures that could permit a subject to obtain the communicated information in normal sensory fashion either inadvertently or through deliberate cheating are particularly fatal. This is called the problem of "sensory leakage." Inadequate procedures for randomizing (randomly selecting) target stimuli are another common problem.

Methodological inadequacies plague all the sciences, but the history of parapsychology is embarrassingly full of promising results that collapsed when the procedures were critically examined (Akers, 1984). One common charge against parapsychology is that preliminary, poorly controlled studies often obtain positive results, but that as soon as better controls and safeguards are introduced, the results weaken or disappear.

Meta-analysis provides an empirical way of evaluating this charge, too. Once a flaw is discovered in a completed experiment, there is no persuasive way of arguing that the flaw did not contribute illegitimately to a positive outcome; the only remedy is to redo the experiment correctly. In a database of several studies, however, it is possible to test whether the more poorly controlled studies in the database did in fact obtain more positive results than the well controlled studies. If there is a correlation between a procedural flaw and positive results across the studies, then there is a problem. In the case of the ganzfeld database, both critic Hyman and parapsychologist Honorton agree that flaws of inadequate security and possible sensory leakage do not correlate with positive results. Hyman claimed to find a correlation between flaws of randomization and positive results, but both Honorton's analysis and two additional analyses by nonparapsychologists dispute his conclusion (Harris & Rosenthal, 1988; Saunders, 1985). Moreover, a series of 10 new studies designed to control for flaws identified in the original database yielded results quite consistent with the original set of 28 studies (Harris & Rosenthal, 1988).

File-Drawer Problem Suppose that each of 20 investigators independently decides to conduct a ganzfeld study. Even if there were no ganzfeld effect, there is a reasonable probability that at least one of these investigators would obtain a significant result by pure chance. That lucky investigator would then publish a report of the experiment, but the other 19 investigators—who obtained "null" results—are likely to become discouraged, put their data into a file drawer, and move onto something more promising. As a result, the scientific community would learn about the one successful study but have no knowledge of the 19 null studies buried in the file drawers. The database of known studies would thus be seriously

biased toward positive studies, and any meta-analysis of that database would arrive at similarly biased conclusions. This is called the file-drawer problem.

The problem is particularly tricky because it is impossible, by definition, to know how many unknown studies are languishing in file drawers somewhere. Nevertheless, parapsychologists offer two lines of defense against the charge that the file-drawer problem seriously compromises their database.

First they point out that the Parapsychological Association may be the only professional scientific society to have an official policy against selective reporting of positive findings and that negative findings are frequently reported in the Association's five affiliated journals. Moreover, the community of parapsychologists is relatively small, and most investigators know about ongoing work in the active laboratories around the world. When conducting meta-analyses, they attempt to ferret out all unpublished studies through their personal networks.

But the major defense is statistical, and again meta-analysis provides an empirical approach to the problem. By knowing the overall statistical significance of the known database, it is possible to compute the number of studies with null results that would have to exist in file drawers to cancel out that significance. In the case of the ganzfeld database, it is estimated that there would have to be over 400 unreported ganzfeld studies with null results—the equivalent of 12,000 sessions—to cancel out the statistical significance of the 28 studies analyzed in the debate (Honorton, 1985). Not surprisingly, there is consensus that the overall significance of the ganzfeld studies cannot reasonably be explained by the file-drawer effect (Hyman & Honorton, 1986).

Rather than continuing their debate, Hyman and Honorton issued a joint communiquè in which they set forth their areas of agreement and disagreement and made a series of suggestions for the conduct of future ganzfeld studies (Hyman & Honorton, 1986). Their debate and the subsequent discussion provide a valuable model for evaluating all disputed domains of scientific inquiry.

Anecdotal Evidence

In the public mind, the evidence for psi consists primarily of personal experiences and anecdotes. Virtually all such evidence is scientifically unacceptable because it suffers fatally from the same problems that threaten experimental evidence—nonreplicability, inadequate controls, and the file-drawer problem.

The replication problem is insurmountable because most such evidence consists of one-time occurrences that cannot be repeated. A woman announces a premonition that she will win the lottery that day—and she does. You dream about an unlikely event that actually occurs a few days later. A "psychic" correctly predicts the assassination of a public figure. And because such incidents occur under unexpected and ambiguously specified conditions, the problem of inadequate controls is also decisive. There is usually no way of ruling out such alternative interpretations as coincidence (chance), faulty memories, and deliberate deception.

The file-drawer problem is also pervasive. The lottery winner who announced ahead of time that she would win is prominently featured in the news. But the thousands of others with similar premonitions who did *not* win are never heard from; they remain in the file drawers. It is true that the probability of this woman's winning the lottery was very low. But the critical criterion in evaluating this case is not the probability that *she* would win but the probability that any *one* of the thousands who thought they would win would do so. That probability is much higher. Moreover, this woman has her own personal file drawer which contains all those past instances in which she had similar premonitions and then lost.

The same reasoning applies to precognitive dreams. We tend to forget our dreams unless and until an event happens to remind us of them. We thus have no way of evaluating how often we might have dreamt of similar unlikely events that did *not* occur. We fill our database with positive instances and unknowingly exclude the negative instances.

The fullest file drawers probably belong to the so-called "psychics" who make annual predictions in the tabloid newspapers. Everybody remembers their occasional "direct hits," but

nobody remembers the predictions that fail. In fact, these psychics are almost always wrong (Frazier, 1987; Tyler, 1977).

Skepticism About Psi

If some of the experimental evidence for psi is as impressive as it seems, why hasn't it become part of established science? Why the continuing skepticism?

Extraordinary Claims Most scientists believe that extraordinary claims require extraordinary proof. A study reporting that students who work harder get higher grades will be believed even if the study was seriously flawed because the data are consistent with our understanding of how the world works. But the claim that two people in a ganzfeld study communicate telepathically is more extraordinary; it violates our a priori beliefs about reality. We thus rightly demand a higher measure of proof from parapsychologists because their claims, if true, would require us to radically revise our model of the world—something we should not undertake lightly. In this way, science is justifiably conservative. Many open-minded nonparapsychologists are genuinely impressed by the ganzfeld studies, for example, but they reasonably can and do ask to see more evidence before committing themselves to the reality of psi.

Extraordinariness is a matter of degree. Telepathy seems less extraordinary to most of us than precognition because we are already familiar with the invisible transmission of information through space. We may not all understand how television pictures get to our living rooms, but we know that they do so. Why should telepathy seem that much more mysterious? Precognition on the other hand seems more extraordinary because we have no familiar phenomena in which information flows backward in time.

Extraordinariness also depends upon our current model of reality. As our understanding of the world changes, a phenomenon that seemed extraordinary at an earlier time may no longer seem so—even if the quality of the evidence has not changed. Any child who has visited a museum of natural history has seen fragments of a meteorite. But before the

asymmetries in the human brain (see Chapter 2) have spawned a host of pop-psychology books and media reports containing unsubstantiated claims about left-brained and right-brained persons. Irresponsible reports about states of consciousness—including hypnosis and psi—appear daily in the media. It is thus pertinent to note that when the college professors in the survey cited above were asked to name the sources for their beliefs about ESP, they most frequently cited reports in newspapers and magazines.

And finally, research in cognitive and social psychology has sensitized psychologists to the biases and shortcomings in our abilities to draw valid inferences from our everyday experiences (see Chapter 19). This makes them particularly skeptical of anecdotal reports of psi where, as we saw above, our judgments are subject to many kinds of errors.

For these several reasons, then, much of the skepticism of psychologists toward psi is well-founded. But some of it is not. As we noted earlier, some scientists declare psi to be an impossibility and reject the legitimacy of parapsychological inquiry—a priori judgments we believe to be out of place in science. Only 4% of the college professors in the survey declared psi to be an impossibility—but 34% of the psychologists did so. Two hundred years ago, these same skeptics would have been equally certain that God does not hurl stones at us from heaven.

Rosenthal, R. (1986). Meta-analytic procedures and the nature of replication: the ganzfeld debate. *Journal of Parapsychology*, **50**, 315-337.

Saunders, D. R. (1985). On Hyman's factor analyses. *Journal of Parapsychology*, **49**, 86-88.

Spence, K. W. (1964). Anxiety (Drive) level and performance in eyelid conditioning. *Psychological Bulletin*, **61**, 129-139.

Stokes, D. M. (1987). Theoretical parapsychology. In S. Krippner (Ed.), *Advances in parapsychological research. Vol. 5* (pp. 77-189). Jefferson, NC: McFarland.

The Steering Committee of the Physicians' Health Study Research Group. (1988). Preliminary report: Findings from the aspirin component of the ongoing Physicians' Health Study. *New England Journal of Medicine*, **318**, 262-264.

Tyler, H. (1977). The unsinkable Jeane Dixon. *The Humanist*, **37**, 6-9

Utts, J. (1986). The ganzfeld debate: a statistician's perspective. *Journal of Parapsychology*, **50**, 393-402.

Wagner, M. W., & Monnet, M. (1979). Attitudes of college professors toward extra-sensory perception. *Zetetic Scholar*, **5**, 7-17.