

How Not to Test Mediums¹

Critiquing the Afterlife Experiments

Professor Gary Schwartz makes revolutionary claims that he has provided competent scientific evidence for survival of consciousness and—even more extraordinary—that mediums can actually communicate with the dead. He is badly mistaken. The research he presents is flawed, and in numerous ways. Probably no other extended program in psychical research deviates so much from accepted norms of scientific methodology as this one.

RAY HYMAN

Gary Schwartz is professor of psychology, medicine, neurology, psychiatry, and surgery at the University of Arizona. After receiving his Ph.D. in personality psychology from Harvard University, he taught at Harvard and then at Yale University for twenty-eight years as a professor of psychology and psychiatry. He has published more than 400 scientific papers. He came to the University of Arizona in 1988 to do research on, among other things, the relationship between love and health. In 1993 he met Linda Russek and married her soon after. Linda was still grieving over the death of her father. Soon after she met Schwartz, Linda asked him, "Do you think it is possible that my father is still alive?"



That question triggered a research program to answer it and the more general question of survival of consciousness. At first the program was conducted in secret and then became public around 1997. Since 1997, Schwartz has reported a number of studies in which he and his coworkers have observed some talented mediums such as John Edward and George Anderson give readings to sitters in his laboratory. This work has attracted considerable attention because of Schwartz's credentials and position. Even more eye-opening is Schwartz's apparent endorsement of the mediums' claims that they are actually communicating with the dead.

For Schwartz this conclusion follows from the famous principle known as Occam's Razor. Schwartz paraphrases Occam's principle as "All things being equal, the simpler hypothesis is usually the correct one."² As Schwartz sees it, "The best experiments [supporting the reality of communicating with the dead] can be explained away, only if one makes a whole series of assumptions. . . ." These assumptions include: 1) that mediums use detectives to gather some of their information; 2) that sitters falsely remember specific facts such as the names of relatives; 3) that the mediums are super guessers; 4) that mediums can interpret subtle cues such as changes in breathing to infer specific details about the sitter and her relatives; and 5) that the mediums use super telepathy to gather facts about the sitter's deceased friends and family. According to Schwartz, such assumptions create unnecessary complexity. "However, if we were to apply Occam's Razor to the total set of data collected over the past hundred years, including the information you have read about in this book, there is a straightforward hypothesis that is elegant in its simplicity. This is the simple hypothesis that consciousness continues after death. This hypothesis accounts for all the data" [p. 254].

Schwartz's new book *The Afterlife Experiments* presents evidence from a series of five reports in which Schwartz and his

associates observed mediums give readings to sitters "in stringently monitored experiments." Schwartz does admit that his experiments were not ideal. For example, only the very last in his sequence of studies used a truly double-blind format. Yet he insists that the mediums, although often wrong, consistently came up with specific facts and names about the sitters' departed friends and relatives that the skeptics have been unable to explain away as fraud, cold reading, or lucky guesses. He provides several examples of such instances throughout the book. These examples demonstrate, he believes, that the readings given by his mediums are clearly different from those given by cold readers and less gifted psychics. "If cold readings are easy to spot by anyone familiar with the techniques, the kinds of readings we have been getting," he asserts, "in our laboratory are quite different in character."

Could It Be Cold Reading?

Now it so happens that I have devoted more than half a century to the study of psychic and cold readings. I have been especially concerned with why such readings can seem so concrete and compelling, even to skeptics. As a way to earn extra income, I began reading palms when I was in my teens. At first, I was skeptical. I thought that people believed in palmistry and other divination procedures because they could easily fit very general statements to their particular situation. To establish credibility with my clients, I read books on palmistry and gave readings

*Ray Hyman is professor emeritus of psychology at the University of Oregon. His lifelong research interests include examination of alleged psychic readings and the psychology of deception and self-deception. Many of his critiques of parapsychological experiments and related matters were collected in his book *The Elusive Quarry* (Prometheus 1989). Address: 5205 Nectar Way, Eugene, OR 97405. E-mail: rayhyman@msn.com.*

according to the accepted interpretations for the lines, shape of the fingers, mounds, and other indicators. I was astonished by the reactions of my clients. My clients consistently praised me for my accuracy even when I told them very specific things about problems with their health and other personal matters. I even would get phone calls from clients telling me that a prediction that I had made for them had come true. Within months of my entry into palm reading, I became a staunch believer in its validity. My conviction was so strong that I convinced my skeptical high school English teacher by giving him readings and arguing with him. I later also convinced the head of the psychology department where I was an undergraduate.

When I was a sophomore, majoring in journalism, a well-known mentalist and trusted friend persuaded me to try an experiment in which I would deliberately read a client's hand opposite to what the signs in her hand indicated. I was shocked to discover that this client insisted that this was the most accurate reading she had ever experienced. As a result, I carried out more experiments with the same outcome. It dawned on me that something important was going on. Whatever it was, it had nothing to do with the lines in the hand. I changed my major from journalism to psychology so that I could learn why not only other people, but also I, could be so badly led astray. My subsequent career has focused on the reasons why cold readings can appear to be so compelling and seemingly specific.

Psychologists have uncovered a number of factors that can make an ambiguous reading seem highly specific, unique, and uncannily accurate. And once the observer or client has been struck with the apparent accuracy of the reading, it becomes virtually impossible to dislodge the belief in the uniqueness and specificity of the reading. Research from many areas demonstrates this finding. The

principles go under such names as the fallacy of personal validation, subjective validation, confirmation bias, belief perseverance, the illusion of invulnerability, compliance, demand characteristics, false uniqueness effect, foot-in-the-door phenomenon, illusory correlation, integrative agreements, self-

reference effect, the principle of individuation, and many, many others. Much of this is facilitated by the illusion of specificity that surrounds language. All language is inherently ambiguous and depends much more than we realize upon the context and nonlinguistic cues to fix its meaning in a given situation.

Again and again, Schwartz argues that the readings given by his star mediums differ greatly from cold readings. He provides samples of readings throughout the book. Although these samples were obviously selected because, in his opinion, they represent mediumship at its best, every one of them strikes me as no different in kind from those of any run-of-the-mill psychic reader and as completely consistent with cold readings. In August 2001, Schwartz assembled a panel of seven experts on cold reading, including me, to instruct him on the topic. We were shown videotapes of Suzane Northrup and John Edward giving readings in his laboratory. Most members of the panel were openly sympathetic to Schwartz's goals and program. Yet we all agreed that what we saw Northrup and Edward doing was no different from what we would expect from any cold reader.

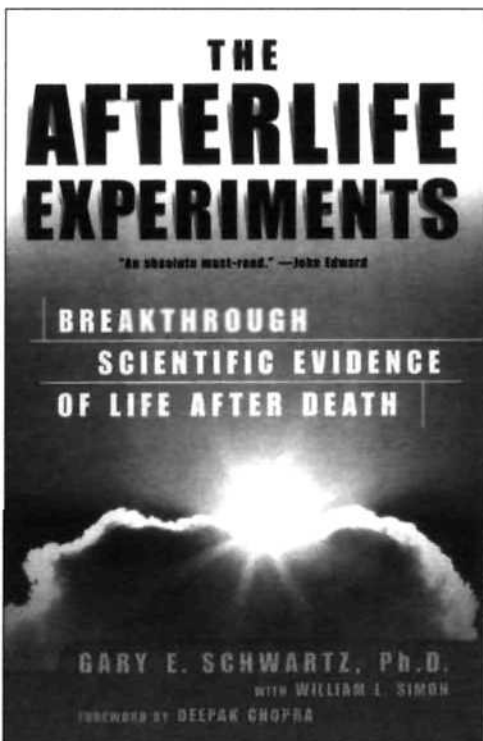
I am sure that Professor Schwartz will strongly disagree with my observation that the readings he presents as strong evidence for his case very much resemble the sorts of readings we would expect from psychic readers in general and cold readers in particular.

This disagreement between us, however, relies on subjective assessment. That is why we have widely accepted scientific methods to settle the issue. That is why it is important, especially for the sort of revolutionary claims that Schwartz wants to make, that it be backed up by competent scientific evidence. Throughout his 2002 book *The Afterlife Experiments*, Schwartz implies that he has already provided such evidence.

This, as I will explain, is badly mistaken. The research he presents is flawed. Probably no other extended program in psychical research deviates so much from accepted norms of scientific methodology as this one does.



Gary E. Schwartz



The Afterlife Experiments: Breakthrough Scientific Evidence of Life After Death. By Gary E. Schwartz, Ph.D. with William L. Simon. Pocket Books, New York, 2002. ISBN 0-7434-3658-X. 374 pp. Hardcover, \$25.

Is the Research Fundamentally Flawed?

Although never going so far as to claim his research methodology is ideal, he apparently believes it is adequate to justify his conclusions that his mediums are communicating with the dead. He writes, "Skeptics who claim that this is some kind of fraud the mediums are working on us have nonetheless been unable to point out any error in our experimental technique to account for the results" (p. xxii). Later he asserts, "The data appear to be real. If there is a fundamental flaw in the totality of the research presented in these pages, the flaw has managed to escape the many experienced scientists who have carefully examined the work to date" (p. 13).

These statements perplex me greatly. I have carefully itemized not one but several "fundamental" flaws in Schwartz's afterlife experiments. I confronted Schwartz with this listing of flaws at two public meetings where we shared the same platform. I also brought them up again at the panel on cold reading that he convened. The other members of the panel also pointed to flaws. And Wiseman and O'Keeffe³ pointed to serious problems with Schwartz's first two published studies in the areas of judging bias, control group biases, and sensory leakage. I would have to make this article almost as long as Schwartz's book to explain adequately each flaw. Because any one of these flaws by itself would suffice to invalidate his experiments as acceptable evidence, I will discuss only a few of these here. First, I will list here the major types of flaws in the experiments described in his first four reports (I will deal with the fifth report separately below):

1. Inappropriate control comparisons
2. Inadequate precautions against fraud and sensory leakage
3. Reliance on non-standardized, untested dependent variables
4. Failure to use double-blind procedures
5. Inadequate "blinding" even in what he calls "single blind" experiments
6. Failure to independently check on facts the sitters endorsed as true
7. Use of plausibility arguments to substitute for actual controls

The preceding list refers to defects in the conduct of the experiments and in the gathering of the data. Other very serious problems appear in the way Schwartz interprets and presents the results of his research. These include:

8. The confusion of exploratory with confirmatory findings
9. The calculation of conditional probabilities that are inappropriate and grossly misleading
10. Creating non-falsifiable outcomes by reinterpreting failures as successes
11. Inflating significance levels by failing to adjust for multiple testing and by treating unplanned comparisons as if they were planned.

Other problems involve failure to use adequate randomization procedures, using only sitters who are predisposed to the survival hypothesis, inappropriate statistical tests, and other common defects that plague new research programs. Even if

the research program were not compromised by these defects, the claims being made would require replication by independent investigators. Perhaps Schwartz's most serious misconception is seen in his attempt to shift the burden of proof from himself to the skeptics.

The worst mistake made by Schwartz and his colleagues was to publish the results they have obtained so far. Instead, they should have first tried to gather evidence for their hypothesis that would meet generally accepted scientific criteria. By submitting their very inadequate studies to public scrutiny and by demanding that skeptics "explain away" their defective data, they have lost credibility. In addition, the journals that did accept these studies for publication and Schwartz's panel of Friendly Devil's Advocates have also suffered greatly in credibility.

The worst mistake made by Schwartz and his colleagues was to publish the results they have obtained so far. Instead, they should have first tried to gather evidence for their hypothesis that would meet generally accepted scientific criteria.

Schwartz's Inadequate and Inappropriate Response to Criticisms

Schwartz's responses to criticisms such as those made by Wiseman and O'Keeffe obscure rather than clarify matters.⁴ For example, regarding his failure to provide safeguards against sensory leakage, he complains that Wiseman and O'Keeffe "curiously did not mention that we were fully cognizant of such issues and were actively researching them at the time the Schwartz et al. paper was published." The fact that the researchers were aware that they had not provided adequate safeguards against sensory leakage does not in any way make their data more acceptable. Indeed, if they were aware of how to properly control for this flaw, it is even more inexcusable that they failed to do so. Why did they publish data they knew to be compromised and try to pass them off as legitimate science?

Indeed, Schwartz actually states that he deliberately allowed for some sensory leakage to see if "the remaining subtle cues" could explain the subsequent accuracy of the mediums' statements. He also states that he wanted to begin with "a semi-naturalistic design . . . to develop a professional relationship with the mediums. . . ." If, in fact, this was his rationale for using an inadequate design, then he should have treated the study as a preliminary probe to see if the mediums could work under laboratory conditions. Such a preliminary or pilot study, however, should then be followed up with a formal, properly conducted experiment. Knowing how to properly control for sensory leakage in no way licenses the publishing of flawed data to support a hypothesis.

In defending himself against the charge of sensory leakage, Schwartz uses another tactic that violates acceptable scientific conduct. He tries to shift the burden of proof onto the

skeptic: "Skeptics who speculate that 'cold reading' can achieve similar results have a responsibility to show that identical findings can be obtained *under the conditions used in the Schwartz et al. research* (e.g., the single-blind sitter-silent condition that effectively rules out pre-experimental information and verbal feedback). We welcome such experiments."

This control group in no way provides a proper comparison or baseline for the "accuracy ratings" of the mediums by the sitters. The control group was given a task that differed in very important ways from that of the mediums.

Sorry, Professor Schwartz. The skeptics and the scientific community have no responsibility to show anything until you provide them with data collected according to well-established and acceptable standards. The responsibility is yours to first provide us with evidence for your hypothesis of survival of consciousness that is gathered according to the appropriate scientific standards which include controlling for sensory leakage; devising dependent variables that are relevant, reliable, and valid; and using control comparisons that are meaningful.

Schwartz's rejoinders to Wiseman and O'Keeffe's other two topics of criticism are even more disturbing. His response to the charge of possible judging bias is that, "The purpose of the original Schwartz et al. experiments (2001) was *not* to rule out possible rater bias, but to minimize it." He again tries to shift the burden of proof to the skeptic, by arguing that it is implausible to speculate that his sitters would exhibit rater bias on such things as names, relationships, and the like. Indeed, it is highly plausible to me that some sitters might acquiesce to statements that are demonstrably false. However, science exists as a way to avoid arguments over plausibility. Minimizing rater bias is not the same as precluding it. If he wants to claim scientific acceptance for his evidence then he has to gather the data under conditions that eliminate or adequately correct for such bias. Even worse is his rejoinder to the claim that he used an inappropriate control group. "The purpose of the original...experiments was *not* to include an ideal control group, but rather to address, and possibly rule out (or in) one possible explanation for the data—i.e., simple guessing."

This last statement is both confusing and wrong. I suspect that Schwartz means by "an ideal control group" one made up of individuals who are the same age and have the same sort of experience as his mediums. Since his actual control group consisted of undergraduate students who had no prior experience as mediums, the group was obviously not ideal in this sense. However, what Wiseman and O'Keeffe are criticizing is that this control group in no way provides a proper comparison or baseline for the "accuracy ratings" of the mediums by the sitters. This is for the simple reason that the control group was given a task that differed in very important ways from that of the medi-

ums. There is no way that the results from this control group could provide a comparison or baseline for simple guessing.

The mediums are free to make statements about possible contacts, names, relations, causes of death, and other matters. In the earlier experiments they were given "yes" and "no" replies from the sitters and in later experiments they typically began a segment without feedback and then went through an additional segment with feedback. The sitters were free to find matches within the output of the medium to fit their particular circumstances. Later the sitter was given a transcript of the entire reading and rated each statement for how accurately it applied to her situation. The statements that got the highest rating were counted as hits. The proportion of such hits varied from approximately 73 to 90 percent in the earlier experiments and somewhat lower in the later ones.

In contrast, the control subjects were given a series of questions based on a reading given to their first sitter. Statements from the readings were converted into questions that could be answered in such a way that the answer could be scored correct or incorrect. For example, if the medium had correctly guessed the cause of the sitter's mother's death, a question given to the controls might be, "What was the cause of her mother's death?" Schwartz and his colleagues report that the average percentage of correct answers by the controls was 36 percent. Because the "accuracy" of the mediums was much higher, the researchers conclude that the mediums had access to true information that cannot be explained away as guessing.

Wiseman and O'Keeffe correctly point out that this is an inappropriate comparison. Although Schwartz claims that, if anything, the controls had an advantage over the mediums, the use of the results for the control groups as a baseline for the mediums is completely meaningless. Wiseman and O'Keeffe provide several reasons why. In addition to the reasons they give, a more fundamental one is that the score for the controls does not involve subjective ratings by the sitters *while the accuracy scores for the mediums depend entirely upon the judgment of these sitters*. We have no idea how well the mediums could do if given the same task as the controls. I strongly suspect they could not perform any better.

The accuracy score for the medium is completely dependent on the subjective decisions of the sitter. The very first example of a reading provided in this book begins as follows:

The first thing being shown to me is a male figure that I would say as being above, that would be to me some type of father image. . . . Showing me the month of May. . . . They're telling me to talk about the Big H—um, the H connection. To me this an H with an N sound. So what they are talking about is Henna, Henry, but there's an HN connection. (p. xix)

The sitter identified this description as applying to her late husband, Henry. His name was Henry, he died in the month of May and was "affectionately referred to as the 'gentle giant.'" The sitter was able to identify other statements by the medium as applying to her deceased spouse.

Note, however, the huge degree of latitude for the sitter to fit such statements to her personal situation. The phrase "some type of father image" can refer to her husband because he was also the father to her children. However, it could also refer to her own father, her grandfather, someone else's father, or any male with children. It could easily refer to someone without children such as a priest or father-like individual—including Santa Claus. It would have been just as good a match if her husband had been born in May, had married in May, had been diagnosed with a life-threatening illness in May, or considered May as his favorite month. The "HN" connection would fit just as well if the sitter's name were Henna or her husband had a dog named Hank.

Schwartz concludes that, "No other person in the sitter's family fit the cluster of facts 'father image, Big H, Henry, month of May' except her late husband, Henry." Of course not! If that person, or any other, also found a match for their personal life, it too would be unique. When I put myself in the shoes of a possible sitter and try to fit the reading to my situation, I can find a good fit to my father, who was physically large, whose last name was Hyman, and for whom, like any human on this planet, experienced one or more notable events in the month of May. Other things in the reading also can easily be fitted to my father. Neither the original sitter nor anyone else would fit this cluster of facts! Schwartz makes much of the fact that the cluster of facts that a sitter extracts from a reading tend to be unique for that sitter. He even calculates the conditional probabilities of such a cluster occurring just by chance. Naturally, these conditional probabilities are extremely low—often with odds of over a trillion-to-one against chance.

The "accuracy" score for the medium, as calculated by the experimenters, depends critically on the sitter's ratings. This allows subjective validation⁵ and uncontrolled rater biases to enter the picture on the side of the mediums. The sitters were deliberately selected because they were already disposed towards the survival hypothesis (that consciousness survives death). Given the statement "some type of father image," the sitter easily fit this to her late husband who was the father of her children. For her, this would get the highest accuracy rating. A more skeptical sitter, realizing the ambiguity in the statement, might give it a lower rating. Given the statement "showing me the month of May," the committed sitter would rate it accurate because her husband actually died in the month of May. A less committed sitter might rate it as less accurate because she realizes that this statement could apply to any significant event that happened to her husband, herself, or her family in May. From the example above, if I were a committed sitter receiving the same reading, I could see myself giving it a score of five out of five (or 100% accuracy) because my father (obviously a type of father image), experienced one or more significant events in May (*showing me the month of May*), was large and overweight and named Hyman (*about the Big H—um, the H connection...an H with an N sound*).

Compare this with the task confronting the control subjects. They would be given a series of questions based on this reading which might go as follows:

1. What was the relation of the deceased to the sitter?
2. What was the name of the sitter's husband?
3. In what month did he die?
4. How was he described by his friends?

The control students would have to come up with the answers *husband, Henry, May, and big* to get a perfect score. The likelihood of anyone, including the mediums, getting all these correct, or even a high percentage of them correct, is very small indeed. It is obvious that this a completely different task from the one performed by the mediums. A strikingly obvious difference is that the sitter's judgments and biases are completely removed from the task given the controls. Indeed, it is just these potential biases and subjective judgments being made by the sitters that obviously cries out for controlling.

Conditional Probabilities

One way that Schwartz assesses the likelihood that his mediums are obtaining their "hits" just by chance guessing is to calculate conditional probabilities of getting a certain pattern of statements that would match the sitter's situation. In the excerpt from the reading I have been using as an example, he might estimate the probability of getting the gender of the sitter's husband as 1/2; the probability of indicating that he was dead as 1/2; the probability of correctly guessing that deceased person was the sitter's husband as, perhaps, 1/6; the probability of guessing the month of death as 1/12; the probability of getting the correct name as 1/15; and the probability that of knowing that he was described by friends as "big" as 1/20 (of course, the particular probabilities being made in most of these cases have to be based on assumptions and guesswork, but Schwartz claims that he errs on the conservative side in making such estimates). The combined probability of correctly getting this particular pattern of matches just by chance would simply be the product of these separate probabilities. In my example, the probability of achieving this particular pattern of matches would be less than 1 out of 86,000.

Such a low probability would seem to clearly rule out chance as an explanation for the results. Most of Schwartz's actual calculations typically lead to probabilities of less than one out of a million or even millions. In one case he calculated the probability that the results could have been obtained by guessing as 1 in 2.6 trillion! If these calculations were appropriate they certainly would clearly rule out guessing as an explanation for the mediums' apparent successes.

Probability, however, is a very slippery concept. Even experts have gone badly astray in trying to apply it to situations in the real world. Some of the reasons why Schwartz's conditional probability calculations are inappropriate and misleading in this context involve highly technical considerations concerning conditional probabilities, independence, sample spaces, and the like. However, you can realize something must be wrong here when you consider that these same types of calculations also provide very low probabilities for any set of matches that any person—the sitter or someone else—finds in a given reading. For example, the pattern of matches that I find in the sample reading with respect to my late father yields

a probability of guessing that is so low as to also rule out chance. And this will be true for any pattern of matches that anyone can find in the same reading. One problem is that Schwartz's calculations do not take into account the enormous variety of possible combinations that could be extracted from a single reading. Each one would be unique to the person for whom that pattern makes sense.

Ironically, such conditional probability calculations could be justified (with some important reservations) for the task given to the control students. Each question they were posed has an explicit answer. If we can make reasonable assumptions about the probability of getting each answer just by chance, and if we can assume that the answers to each question are independent of each other, then we might legitimately try to estimate the probability of getting all the answers correct by multiplying together the probabilities of correct answers for all the questions. Notice that we can do this only because we defined the *total set of possibilities* and have not selected, after the fact, just those questions that were answered correctly.

Reliance on Uncorroborated Sitter Ratings

This discussion of the reasons why the control comparison and the calculation of conditional probabilities are inappropriate points to one of the most serious weaknesses in this research program. The "accuracy" ratings of the mediums depend entirely upon the judgments of the individual sitters. Each sitter is solely responsible for validating the reading given to him or her. Each sitter is carefully chosen to be someone who is favorably disposed to the survival hypothesis and who wants the medium to be able to communicate with their departed family and friends. Schwartz admits that the "accuracy" ratings from sitters who are not so favorably disposed are much lower. Although this is consistent with rater bias, Schwartz has other explanations. He also believes that just as some mediums are "white crows," there are also sitters who are "white crows"—that is, some sitters are prone to get especially good results. In other words, some sitters are more prone to give higher ratings of accuracy than do other sitters.

One simple explanation, consistent with Occam's Razor, is that some sitters are more susceptible to response biases. Schwartz, I am sure, will strongly disagree. This, again, highlights the need for properly conducted research that precludes

or adequately corrects for such possible biases. This is why a properly conducted research program requires carefully standardized, reliable, and valid dependent variables; truly double-blind procedures; appropriate control comparisons; and proper controls for sensory leakage. All of these requirements, as I have explained, are lacking in the afterlife experiments.

Schwartz has tried to counter some of these criticisms by pointing to the fact that much of the information provided by the medium consists of factual material that can be independently checked (for example, specific names,

relationships, careers, gender, etc.). Yet he has never bothered to make an independent check on these "facts." He simply accepts the sitters' statements. He argues that it is completely unreasonable to believe that one of his trusted sitters would say "yes" to a fact that was untrue. This, of course, is using a plausibility argument in the place of a control that should have been incorporated into the research. Perhaps it is unlikely that a sitter would acquiesce to a factual statement that she or he knows to be untrue. However, his own excerpts from readings given in his book provide one or more examples. In one case, one of his best sitters keeps acquiescing to John Edward's mistaken belief that her husband is dead, even though he is alive and sitting in the next room. As he does over and over again when he encounters what looks like a miss, Schwartz manages to find a convenient explanation to this peculiar situation. He suggests that this could be case of precognition because the sitter's husband was killed in an accident some months after the reading.

The Laurie Campbell "White Crow" Readings

The book begins with a quotation from William James. "In order to dis-

prove the law that all crows are black, it is enough to find one white crow." James was interested in the possibility of psychic phenomena. He believed that it was sufficient to find one truly indisputable example of a psychic occurrence to demonstrate that violations of natural law were possible. Schwartz claims he has uncovered several white crows. The performance of his mediums, especially Laurie Campbell and John Edward, earn them the accolade, in his judgment, of "white crow" mediums. He has also found at least one "white crow" sitter in one of his participants, GD.

GD is a psychiatric social worker who lost his partner,



Gary Schwartz examines data from his experiments. Frames from *Dateline NBC*.



One of the tested mediums, left, tries to get information from the sitter.



John Edward is tested by Gary Schwartz.

Michael, to AIDS. GD discovered he had mediumistic powers and believed he was in contact with his deceased partner. He took part as one of three sitters in an experiment with the medium Laurie Campbell. The researchers reported that, "Statistically significant evidence for anomalous information retrieval was found for each of the three sitters investigated in this experiment. However, it is the uniqueness and extraordinarily evidential nature of the particular reading highlighted in this detailed report that justifies focusing on this 'white crow' research reading." In other words, the researchers base their report entirely on the results with this one sitter. Although one of the criteria for the selection of the sitters was their willingness to rate the transcripts of their readings, such ratings were apparently not done at the time this report was written. The experimenters report that GD estimated that the information given by the medium was at least 90 percent accurate. Presumably this was simply a subjective estimate. In the previous experiments the "accuracy" rating was obtained by calculating the proportion of highly rated items among all of the rated items.

Schwartz et al. state that the complete reading took over an hour. They promised that the full transcript will be made available at some future date. So far, I have not seen it, so I cannot judge to what extent this reading might be qualitatively different from the readings that I have witnessed or read that have been given by Laurie Campbell. In the readings I am familiar with, Campbell throws out initials, names, and vague statements that appear to me to characterize the readings from the many psychic readers and mediums I have studied over the past sixty years. I witnessed a public demonstration by her at a conference sponsored by Gary Schwartz and Linda Russek in Tucson in March 2001. I have also carefully studied the complete transcripts of two readings by Campbell.

At first blush the reading given for GD appears qualitatively different. From what we are told, Campbell apparently stated that the recipient of the reading was named George (true) even though she was supposedly completely blind to his identity. She also correctly indicated that the primary deceased person for GD was a male named Michael (true). She also provided the name "Alice" and later, during the interactive part of the reading, correctly stated that this was GD's deceased aunt. Among the list of names she included in her reading was one that she said sounded like *Talya*, *Tiya*, or *Tilya*. GD has a friend that he calls "Tallia." Campbell mentioned a deceased dog whose name began with an "S." GD had a beloved dog with an "S" name (but not the name used by Campbell). Other names were also relevant including that of GD's father "Bob." The researchers cite other qualitative hits that they believe provide powerful evidence that Campbell is getting information from a paranormal source.

This paranormal source, the authors argue, is not simply extrasensory perception based on GD's thoughts. This is because in the interactive phase of the reading "not only were

each of the *four primary people described accurately by Campbell, but four additional facts not known by GD* and later confirmed by sources close to GD indicated that exceptionally accurate information was obtained for GD's deceased and close friends." Because of this, Schwartz argues that the medium is most likely getting her information from the deceased individuals rather than from the sitter's thoughts. At the time of the reading, GD mistakenly thought that Campbell had erred by stating that the granddaughter of his aunt Alice was named "Katherine" because he believed the name was spelled "Catherine." When

The medium throws out initials, names, and vague statements that appear to me to characterize the readings from the many psychic readers and mediums I have studied over the past sixty years.

GD later checked, he discovered that his cousin's name was indeed spelled with a K instead of the C that he was thinking during the reading. Another striking example is where Campbell said "that M [Michael] showed her where he lived; somewhere in Europe, and his parents have a 'heavy accent' (M was German). Laurie Campbell reported that M was showing her a big city, and then M was traveling through the countryside to his home. . . . Campbell claimed that M showed her an old, stone 'monastery' on the edge of the river on the way to his parent's home. This information was not known to GD prior to the reading. After the reading, GD telephoned M's parents in Germany and learned that there was an old abbey church along the river's edge on the way to their house, and that they had held a service for M in this monastery-like stone building a few weeks prior to the experiment."

These are examples from this reading that Schwartz insists that the skeptics cannot explain away in terms of normal causes such as guessing and cold reading, fraud, or unwitting sensory leakage. However, the experiment is compromised by so many serious defects that it would be futile for a skeptic to accept this challenge. This would be another example of placing the burden of proof on the wrong shoulders. Although the experimenters try to make a plausible argument against collusion between Campbell and GD, as well as against the possibility that Campbell might somehow have gotten access to the manuscript of GD's forthcoming book (a copy of which was in Schwartz's) possession, the actual controls against such sensory leakage were not very convincing. Indeed, the authors partially acknowledge this defect. "Since the exceptional nature of the data reported here was not anticipated ahead of time, the experiment did not include additional desirable controls. . . ." Although I see no reason to assume that fraud did occur in this instance, I believe that the experimenters have an obligation to their mediums and sitters, as well as to the scientific community, to take all reasonable steps to preclude

fraud as a possibility. By taking such steps they protect their subjects from any suspicions that might arise in this area.

The results would have become more interesting if they had been collected under double-blind conditions—that is, under conditions where Campbell, GD, and the experimenter, Schwartz, were all in ignorance of one another at the time of the reading. Schwartz calls the experiment “single-blind” because at the time of the reading (at least the first portions of it), GD did not know who the medium was and Campbell did not know who the sitter was and was separated from him by a thousand miles. Unfortunately, the experimenter, who *did* know the identity of the sitter as well as quite a bit of his personal history, was with Campbell at the time she was giving much of the reading. Psychical researchers have a long history of dismissing data collected with this weakness as non-evidential.

Probably the most serious weakness of this experiment is that its outcome relies entirely upon the uncorroborated judgments of the sitter GD. Again, Schwartz relies on plausibility arguments for the reliability and validity of GD's ratings of the reading. This is a major defect for many reasons. One is simple rater bias. Individuals can differ widely as to what they will or will not accept as valid for their personal situation. When Campbell says that she is hearing a name that sounds like *Talya*, *Tily*, or *Tilya*, a sitter with a strict criterion might not accept this as referring to a friend whose name is *Tallia*. On the other hand, a sitter with a looser criterion and who is convinced that the medium is talking about his situation might accept Campbell's probe as referring to a friend with the name of *Tanya*, *Tina*, *Tilda*, *Tony*, *Dalia*, *Natalie*, or a variety of other possibilities. Schwartz may be right that it is unlikely that GD would misremember or misreport having a friend by the name of *Tallia*. However, if the outcome of this reading is so earth-shaking and scientifically revolutionary as he claims it is, I would think that he should at least make the effort to independently check on some of these facts.

This is especially true for “facts” that were unknown to GD at the time of the reading, but were later discovered by him to be true. For example, when GD called M's parents in Germany, how did the questioning take place? Did they speak in German or English? How well does GD speak German? How well do M's parents speak and understand English? Did GD ask the questions in a leading way? Certainly it would have been highly desirable for the experimenters to have independently communicated with the M's parents. Indeed, it would have been better if they, rather than GD, did all the checking. Instead, everything depends upon GD. Such reliance on a single individual in such circumstances is called by psychologists “the fallacy of personal validation.”

“Replication” of the Laurie Campbell/GD Reading in a Double-Blind Experiment

What is required, of course, is a successful replication of these apparently spectacular results in a reading conducted under properly double-blind conditions. Indeed, this is precisely what Schwartz claims he has achieved. He and his colleagues finally conducted a double-blind experiment using Campbell as the medium and six sitters, one of whom was GD. During

the readings, Campbell and the sitters had no contact and the two experimenters who were with Campbell were blind to the order in which the sitters were run. Later each sitter was sent two transcripts to judge. One was of the actual reading for that sitter and the other was of a reading given to another subject. The sitters were given no clues as to which was their actual reading. “The question was, even under blind conditions, could the sitters determine which of the readings was theirs?”

The findings were breathtaking. Once again it was George Dalzell's [GD's] reading [that] stood out. . . . This provided incontrovertible evidence in response to the skeptics' highly implausible argument against the single-blind study that the sitter would be biased in his or her ratings (for example, misreading his deceased loved ones' names and relationships) because he knew the information was from his own reading. . . . The skeptics' complaint becomes a completely and convincingly impossible argument in the case of the double-blind study. . . . It appeared to be the ultimate “white crow” design. . . . (p. 236)

As these quotations reveal, Schwartz believes this double-blind experiment has put to rest all the skeptical arguments against his evidence. One of Schwartz's mantras in relation to his afterlife experiments is *let the data speak*. When I read the full the report of this “ultimate ‘white crow’ design,” the data did speak loud and clear. However, the story the data told is just the opposite from the one that Professor Schwartz apparently hears.

The plan of the study was admirably simple. Campbell gave readings to the six sitters in an order that neither she nor the experimenter who was with her knew. In this way neither the medium nor the person in her presence was aware of who the sitter was at the time of the reading.⁷ At the time of the reading, the sitter was physically separated from the medium. The medium gave her readings in Tucson, Arizona, while the sitters were in their homes in different parts of the country. Subsequently, each sitter was mailed two transcripts. One of the transcripts was the actual reading for that sitter and the other was from the reading of another sitter. Each sitter rated the two transcripts, not knowing which was the one actually intended for her or him, according to instructions provided by the researchers. The sitter first circled every item in the transcripts which they judged to be a “dazzle shot.” “For you, a dazzle shot is some piece of information—whatever it is *to you*, that you experience as ‘right on’ or ‘wow’ or ‘that’s my family.’” Next, the sitter was instructed to go through the transcripts again and score each item as a hit, a miss, or unsure.

Finally, the sitter designated which of the two transcripts was the one that actually was intended for him or her.

The hypothesis was that if Campbell could truly access information from the sitter's departed acquaintances, this would show up on all three measures. In other words, the sitters would successfully pick their own reading from the two transcripts; they would record significantly more dazzle shots in their own transcripts as compared with the control transcripts; and they would find many more hits and fewer misses in the actual as opposed to the control transcript. *Each one of these three predictions failed*. Four of the sitters did correctly pick their own transcript, but this is consistent with the chance expectation of three successes. On the two more sensitive

measures, there were no significant differences in number of dazzle shots or hits and misses.

The authors admit that for the overall data, "there was no apparent evidence of a reliable anomalous information retrieval effect." So how can they use these results to proclaim a "breathtaking" vindication of their previous findings? This is because, when they looked at the results separately for each sitter, they discovered that in the case of GD, who had been the star sitter in a previous experiment with Campbell, he not only successfully identified his own transcript but also found nine dazzle shots in this transcript and none in the control. The results for the hits and misses were equally striking. He found only a few misses in his own transcript and a large number of misses in the control. He found many hits in his own transcript and not a single one in the control transcript. Given this "unanticipated replication," the authors hail the results as compelling support for their survival hypothesis. However, for anyone trained in statistical inference and experimental methodology, this will appear as just another blatant attempt to snatch victory out of the jaws of defeat. An accepted principle of research methodology is that the reporting of statistical significance from experimental findings derives meaning from the fact that the experimenter specifies *in advance* which comparisons he or she will test. If the experimenter plans to make many comparisons, then the criteria for statistical significance must be adjusted to take into account that the more comparisons that will be made the more chances there will be to find something "significant" just by chance. In the present case, it was obvious that the planned comparisons involved the overall differences between the ratings of the actual and the control transcripts. The authors do not indicate whether they intended to make adjustments for the fact that they were using three different measures, but, in any case, it does not matter because there were no meaningful differences on any of the three indicators.

Of course, these strictures do not preclude the investigators from noticing unexpected outcomes in their data. Such unplanned outcomes can serve as hypotheses for new experiments. When an experimenter finds unanticipated, but interesting, quirks in the data, he or she cannot draw conclusions until the surprise finding has been cross-validated with new data. The reason for this is simple. Any set of data that is reasonably complex will always, just by chance, display peculiarities. Some statisticians and methodologists do allow testing for unexpected findings by means of "post hoc" tests. Such tests require that the departures be much greater than those needed for planned comparisons before they can be declared "significant." Furthermore, such post hoc tests on specific subparts of the data are typically licensed only when the overall tests are significant, which is not the case for the present situation.

So, by commonly accepted scientific practice, the experiment has failed to support the hypothesis it was planned to test. Furthermore, because nothing significant was found, the

results do not warrant claiming a successful replication of previous findings. For scientific purposes, this is all that need be said. However, it may be edifying to discuss some additional reasons why the claim for a successful "replication" is highly suspect in the present case. Three of the six sitters for this experiment were selected just because Campbell had provided "successful" readings for them in previous experiments. They were included to see if she could do so again. For two of them,

**The authors admit that for the overall data,
"there was no apparent evidence of a reliable
anomalous information retrieval effect."
So how can they use these results to proclaim
a "breathtaking" vindication of their
previous findings?**

the authors admit that she failed. So it is only for GD that, in their view, she apparently succeeded.

Comparing the two readings that Campbell gave GD, I find little to support the claim that the second one replicates the apparent success of the first one. Although a full transcript of the first GD reading is still not available, what was included in the first report strongly suggests that the second reading cannot be considered to be aimed at the same individual for whom the first one was given. GD's major interest in mediumship is to establish contact with his deceased partner Michael. Campbell is given credit in the first reading for stating that there was a deceased friend named Michael and then later that he was the primary person for this sitter. The name Michael or a deceased partner does not come up in the second reading. Ironically, the name Michael does appear in the control reading. In the first reading Laurie Campbell mentions a strange name that sounded like *Talya*, *Tiya*, or *Tilya*. GD stated that he indeed had a friend (living) named Tallia. No such name appears in the second reading. Indeed, of the twenty names Campbell produced in the first reading only three come up in the second reading, and these are such common ones as *George*, *Robert* or *Bob*, and *Joe* or *Joseph*. In none of these three cases does she identify whether the person is living or dead or what relationship he has to GD. None of the "specific" facts that she apparently stated during the first reading come up in the second one.

Schwartz claims that the rater bias could not have affected the ratings of this double-blind experiment. A look at GD's dazzle shots and his discussion of the hit and miss data suggests otherwise. His first dazzle shot is "Bob or Robert." These names occur early in the reading in a statement that goes, "And then I could feel like what I thought was like a divine presence and the feeling of a name Mary or Bob or Robert." This appears in a context with other names and other general statements, none of which even hint of a father. The second dazzle shot is "George." Again this appears in a context with no hint

that this could be referring to the sitter. Campbell states, "I got like some names like a Lynn, or Kristie, a George." His third dazzle shot is the statement, "I had the feeling of a presence of an Aunt." GD identifies this aunt as his aunt Alice, although Campbell does not provide the name Alice anywhere in the reading. I count at least twenty-seven names thrown out by Campbell during this second reading. Actually, she covers a much broader range of names because she typically casts a wide net with statements like: "And an 'M' name. More like a Margaret, or Martha, or Marie, something with an 'M.'" It is up to the sitter to find a match. As indicated by his dazzle shots, GD is strongly disposed to do so.

In his qualitative commentary, GD was obviously influenced in selecting one of the transcripts as his reading because it begins with the statement, "I kept feeling the presence of a male." The control reading happens to begin with the statement, "Now, um, to start with I felt like a woman's energy." GD wrote, "I was impressed that the reading is gender specific and accurate. . . ." Instead of assuming that Campbell was somehow conveying information to GD from his departed relatives, it is just as plausible to assume that once GD decided that the actual transcript was meant for him, then subjective validation took over and did the rest. There is, of course, a 50/50 chance that the actual reading is the one that GD will decide is meant for him. From then on, he would read that transcript as if it were truly describing his departed relatives and reject the other as not relevant.

This conjecture fits well with everything we know about subjective validation and the acceptance of personality sketches that one believes was meant for one's self. Is this far-fetched in GD's case? To me, it seems quite obvious just reading the transcript and looking at GD's ratings. The entire case for the reading's validity is based on the assumption that Campbell is describing GD's summer vacation home on Lake Erie in upstate New York. Given this assumption everything is then interpreted within this context. Of course, Campbell never states that she is describing a summer vacation home. It is GD who makes this connection. As just one of many examples of how GD is creative in making the reading fit his circumstances, he gives Campbell credit for having identified the color of their summer cottage which was painted yellow with white trim on the windows. Campbell does, at one point, say, "And I kept getting colors of like yellow and white." This is in a context where she is talking about a woman who spends all her time in the kitchen. One could construe this as perhaps describing the interior colors of the kitchen, the woman's clothing, the old mixer she is described as using, among other possibilities. However, the statement is far removed for any mention of the exterior of the house as such. Earlier in the reading she mentions a white house. A little bit further on, she again mentions a house. She immediately follows this with "And I kept seeing the colors of like grays and blues, but that looked real weathered." Obviously, if the house had been gray and blue, Campbell would have been given credit for a direct hit. GD manages to ignore this and gives Campbell credit for having correctly described the house as yellow and white.

Again, I suspect that Schwartz will disagree with my interpretation. After all, he has already gone on record that this study "provided incontrovertible evidence in response to the skeptics' highly implausible argument against the single-blind study that the sitter would be biased in his or her ratings (for example, misrating his deceased loved ones' names and relationships) because he knew that this information was from his own reading." Nevertheless, the data are quite consistent with the possibility that all we have to do to account for his "breathtaking" findings is to assume that they are due to rater bias.

Conclusions

So what is the bottom line? *The Afterlife Experiments* describes a program of experiments described in four reports using mediums and sitters. The studies were methodologically defective in a number of important ways, not the least of which was that they were not double-blind. Despite these defects, the authors of the reports claim that their mediums were accessing information by paranormal means and that the application of Occam's Razor leads to the conclusion that the mediums are indeed in contact with the departed friends and relatives of the sitters. Schwartz's demand that the skeptics provide an alternative explanation to their results is clearly unwarranted because of the lack of scientifically acceptable evidence. A fifth report describes a study that was designed to be a true double-blind experiment. The outcome, by any accepted statistical and methodological standard, failed to support the hypothesis of the survival of consciousness. Yet the experimenters offer the results as a "breathtaking" validation of their claims about the existence of the afterlife. This is another unfortunate example of trying to snatch victory from the jaws of defeat.

Notes

1. Fans of Martin Gardner will recognize the similarity of this title to that of Martin's book *How Not to Test a Psychic* (1989, Prometheus Books). I thank Martin Gardner for his agreeing to let me adapt his title for this review.
2. The principle usually attributed to William of Occam is typically stated as "entities are not to be multiplied beyond necessity." This statement, as such, cannot be found in the extant writings of William. The principle was known before William was born. However, he did write many different statements that are consistent with the principle such as, "It is vain to do with more what can be done with fewer."
3. Wiseman, R., and C. O'Keeffe. 2001. Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A critique. *The Paranormal Review*, 19: 3-6. (Also in the *SKEPTICAL INQUIRER*, November/December 2001.)
4. Schwartz, G.E. 2001. Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A call for balanced evidence-based skepticism. *The Paranormal Review*: 20.
5. For discussion of this concept and for a very striking illustration of subjective validation in operation see Marks, D. (2000, second edition), *The Psychology of the Psychic*. Amherst, N.Y.: Prometheus Books.
6. Schwartz, G.E., S. Geoffrion, J. Shamini, S. Lewis, and L. Russek. (Submitted to the *Journal of the Society for Psychical Research*.) Evidence of anomalous information retrieval between two research mediums: Replication in a double-blind design. (I obtained a copy of this report from Professor Schwartz in August 2001.)
7. Unfortunately, the double-blind procedure was not ideal. The research coordinator, who was aware of the sitter's identity, phoned Laurie Campbell and the sitter just before the reading. In this way, the medium had contact with someone who was aware of sitter's identity just prior to the reading. □