

BOOK REVIEWS

THE AFTERLIFE EXPERIMENTS: BREAKTHROUGH SCIENTIFIC EVIDENCE OF LIFE AFTER DEATH by Gary E. Schwartz with William L. Simon. New York: Atria Books, 2002. Pp. xxiv + 376. \$13.00 (paperback). ISBN 0-7434-3659-8.

Gary Schwartz, author of *The Afterlife Experiments*, has an impressive set of academic credentials. After receiving his PhD in psychology from Harvard, he moved to Yale, where he served for 28 years as a professor of psychology and psychiatry, director of the Yale Psychophysiology Center, and codirector of the Yale Behavioral Medicine Clinic. In 1988, he moved to the University of Arizona, where he is a professor of psychology, medicine, neurology, psychiatry, and surgery. He has published more than 400 scientific papers.

In this book, Schwartz reports several experiments in which mediums attempt to provide "sitters" with information allegedly obtained from their deceased friends and relatives. He argues that the experiments provide compelling evidence for the "afterlife" hypothesis: that deceased persons can communicate with the living through human mediums. The studies have provoked strong methodological criticisms from Wiseman and O'Keeffe (2001) and from Hyman (2003). In turn, Schwartz has replied to those critiques (2003) and posted an expanded version of his rejoinder on his website (<http://www.openmindsciences.com>).

As in most research programs, the earlier experiments are less formal and less well controlled than subsequent ones in which earlier flaws are corrected and stricter safeguards are introduced. Accordingly, drawing the line between pilot experiments and formal experiments can be problematic. Schwartz himself does not draw such a line and argues for considering the cumulative results, including those obtained in experiments with serious flaws, as evidence for the afterlife hypothesis. In contrast, Hyman (2003) argues that because even the later experiments contain some methodological flaws, all the experiments should be regarded as preliminary and should not have been published at this point in the research program.

The book is cowritten with a professional writer and, as is often the case with books intended for a general audience, the tone is chatty and full of breathless excitement over the accuracy of the mediums. Much of the text is devoted to verbatim quotes of their statements, accompanied by the sitters' (and Schwartz's) subjective evaluations of how well the statements correspond to known facts about the deceased. These correspondences provide the book's "wow" factor, frequently

dazzling the sitters, the investigators, and, presumably, readers with their amazing precision.

The problem with such subjective validations, of course, is that sitters—who are typically highly motivated believers in the afterlife hypothesis—are free to selectively interpret ambiguous or multiple-choice statements (e.g., “I am getting a name that starts with an ‘M’ or an ‘R.’”) to fit any of several of their deceased friends or relatives. In fact, many of the sitters were selected precisely because they had suffered several recent deaths among their acquaintances. Nevertheless, there are rigorous ways of dealing with this problem, and this review will focus on the experimental controls and the quantitative treatment of the data. The informal style of the book often makes it difficult to extract the necessary information, but some of the missing details can be found in the appendixes, where the published reports of some of the experiments are reproduced.

Three questions are at issue: Does the quantitative treatment of the data reveal significant medium accuracy? If so, are the methodological controls adequate to rule out artifactual sources of that accuracy? If the mediums do display significant, nonartifactual accuracy, can it be attributed to communications with the deceased or should it be attributed to psi among the living?

The first experiment reported employed two mediums, Susy Smith (SS) and Laurie Campbell (LC), who had never met each other. Four deceased persons, including William James and personal acquaintances of the participants, were selected. Working alone in her own home, SS “contacted” the deceased persons and asked each to suggest a picture she could draw on their behalf. She then drew four pictures, each associated with one of the deceased, and one control picture that she created for herself. She sealed the pictures in an envelope.

Subsequently, LC participated in two videotaped sessions in which she attempted to contact each of the deceased individuals and obtain descriptions from them of the pictures they had suggested to SS. She also attempted to describe the control picture. After these sessions were recorded, SS brought in the sealed envelope containing the pictures, meeting LC for the first time. The five pictures were removed from the envelope, and LC herself, Schwartz, and two other participants independently served as judges, attempting to match LC’s descriptions to the pictures. Schwartz emphasizes that this study is double-blind: SS drew the pictures before meeting with LC or the experimenters, and neither LC nor the experimenters knew the content or the identity of the pictures until after LC’s descriptions were recorded.

Each of the four judges correctly matched each of the five descriptions to the five pictures, and Schwartz reports that “the combined probability of getting five out of five in four tries is less than one in a thousand (p. 40).” This statement, however, is incorrect. He is here committing the well-known “stacking” error by treating each of the four judges’ matchings

as an independent test. This is actually a one-trial experiment in which the probability of matching all 5 pictures correctly is $1/120$ —no matter how many judges contributed to the matching. The high interjudge agreement does suggest that LC's descriptions were sufficiently unambiguous to be reliably matched, but the number of judges does not affect the significance level. To see this, suppose that SS had explicitly numbered the pictures from 1 to 5 and that LC had correctly stated each of the corresponding numbers in her descriptions. The matching task now becomes trivial, and it should be clear that piling up ("stacking") more judges would not increase the statistical significance of the resulting perfect matching. I mention this statistical error not because it is of major importance in evaluating the study—the result is statistically significant by conventional standards—but because it is representative of several other troubling errors in the book that one would not expect from someone with Schwartz's background.

In this experiment and throughout the book, Schwartz argues that the results favor the afterlife hypothesis rather than the alternative explanation that the information is being obtained through psi among the living. But the data in this study actually favor the psi interpretation. For example, SS was asked to draw a control picture because "if [LC] correctly identified the control picture, this might imply that she was receiving the information through 'remote viewing' of [SS's] apartment, reading [SS's] mind long distance" (p. 39). In fact, the judges easily identified the control picture as matching LC's description, and "even LC was shocked by the clarity of her vision of the control picture" (p. 40). Moreover, LC accurately described the interior of SS's apartment, down to the placement of the furniture and the paintings on the wall.

To pit the afterlife hypothesis against the psi hypothesis, Schwartz conducted a follow-up experiment in which he created a list of 12 people, 6 living and 6 dead, writing the name of each on a separate index card. As LC sat in the same room with him, he pulled out an index card, concentrated on the person listed, and then asked her three questions: Is the person male or female? Young or old? Living or dead? Schwartz hypothesized that "the living subjects would not be aware of LC's attempts to communicate with them, and therefore they would be unlikely to 'communicate' with her. So we expected she would receive more information from the deceased people than from the living" (p. 43). Even though LC did not know who was on the list of names, she correctly answered all 3 questions for all 12 people, a perfect score. If we ignore the problems of potential sensory leakage and concerns about randomization using a closed deck of alternatives, then this result favors the psi hypothesis. In fact, the most parsimonious interpretation of this study is that it is a (badly controlled) test of telepathy between the medium and the experimenter.

The first large-scale experiment reported in the book was funded by a television production company preparing a documentary for HBO on the survival of consciousness. It also involved Laurie Campbell (LC) in addition

to four nationally known mediums whose task was to contact deceased acquaintances of two sitters previously unknown to them. Unfortunately, the requirements of HBO filming and the desire to permit the mediums to work in ways comfortable to them badly compromised the methodological requirements for a sound experiment.

The mediums and sitters were not permitted to see one another before or during the sessions. In each session, one of the mediums and one of the sitters sat in chairs next to each other facing the camera with a partition between them. The medium was permitted to conduct the session in his or her own usual style except that the sitter was permitted to answer questions only with one-word yes/no responses. Obviously this arrangement fails to control for sensory cues. The sitter's voice reveals his or her sex and approximate age, and the tone of the responses can reveal emotional reactions to the questions, which, in turn, can guide the medium's subsequent statements. One need not accuse the mediums of fraud or deliberate use of "cold reading" techniques to appreciate that this procedure is woefully inadequate.

The first sitter received readings from all five mediums, but because of time constraints, the second sitter received readings from only two mediums. In his initial analysis of the sessions, Schwartz attempts to assess the accuracy of the mediums' statements by estimating the probabilities of obtaining the correct answers by chance. For example, all five mediums had reported information about a deceased son of the first sitter:

This is like flipping a coin and getting five heads in a row. None reported receiving any information about a deceased daughter. Again, correct: there was no deceased daughter—three daughters, but all living. So, again, the equivalent of getting five more heads. The probability of getting just this single string of 10 hits (five mediums reporting a dead son and none of them reporting a dead daughter) is approximately one in a thousand by chance. (p. 113)

There are several problems with this statement, but perhaps the most striking is counting as "hits" the fact that the mediums did *not* mention a daughter. Presumably they also did not mention fig trees, or unicorns, or....

Schwartz continues by noting that three of the mediums mentioned the (correct) initial "M" for the sitter's son's name. None reported any other initial. To be consistent with his counting the omission of daughters as a hit, shouldn't he here count the omission of the 25 other letters of the alphabet as hits? Apparently even he sees the absurdity of that. Instead, he estimates that there are at least 16 possible initials that can reflect common first names of males and calculates that "the probability of three mediums

getting the same correct initial is 16 times 16 times 16 (= 4096), which is less than one in four thousand by chance" (p. 114). (Presumably he means that the probability is the *reciprocal* of 4096.) Next, he combines these two sets of observations, noting that the probability of all five mediums getting a deceased son, none guessing a deceased daughter, and three mediums getting the correct initial would be (the reciprocal of) 1024 times 4096, concluding that this is "less than one in four million by chance" (p. 114). He continues to compound the probabilities in this way, arriving at extremely low probabilities that the mediums were responding by chance.

Even as rough estimates, these calculations are—dare I say—not kosher. First, Schwartz again commits the stacking error by treating the number of mediums rather than the number of independent trials as the relevant datum. To assess a medium's accuracy, he or she would have to be tested across a series of independent trials in which, say, the boy/girl alternatives were randomly sampled. To assess all the mediums, each medium would have to be tested on a *different, independent* sequence of trials. In this experiment, there is only one trial—Is this particular deceased child a boy or a girl?—and no matter how many mediums participate, the probability that they, as a group, will correctly identify the sex of the deceased child remains approximately .5. To repeat the point, probability levels are determined by the number of *independent* trials, not the number of judges stacked onto a single trial.

Second, combining probabilities by multiplying them assumes that the events are themselves independent, not conditionally related. For example, suppose that a medium correctly identifies a deceased child as a boy and also notes correctly that he played football. Because being a boy and playing football are positively correlated variables, one cannot compute the combined probability that the child is a boy *and* plays football by simply multiplying the two probabilities.

The third and most serious error in this analysis is that the statements Schwartz selects for his probability analysis are just that: selected. In any given reading, there are many statements that are incorrect and a huge pool of facts about the sitter's deceased acquaintances that are not even mentioned in the reading, but neither kind of "miss" enters into Schwartz's calculations. To calculate such probabilities, one must define the total set of possibilities ahead of time, which is unknowable in this case. One cannot validly select, after the fact, just those questions that were answered correctly. Don't try that at a casino!

In fairness to Schwartz, it should be acknowledged that he treats this probability analysis only as an initial, informal look at the results. His narrative would actually have been more persuasive had he simply omitted this bogus analysis and stuck with a more honest and straightforward "wow, look at this!"

In the subsequent, more formal quantitative analysis of the data, all the statements of the mediums were transcribed and sorted into categories

(e.g., historical fact, initials or names, personal descriptions). The two sitters returned to the laboratory and assigned a rating to each statement made by each of the mediums, ranging from -3 ("complete miss") to +3 ("definite hit"). One sitter assigned a +3 rating to 83% of the statements; the other assigned that rating to 77% of the statements.

In an attempt to assess whether such hit rates could have been obtained by pure guessing, Schwartz selected 70 statements from the readings and turned them into questions. Some were simple yes/no questions (e.g., Is her son dead?); others were more specific (e.g., What was the cause of the child's death?). This questionnaire was administered to 68 students at the University of Arizona. They obtained a mean hit rate of 36%, significantly lower than either of the hit rates observed in the sitters' ratings.

The major problem with this baseline control should be apparent: The task given to the students was very different from that confronting the mediums and much more difficult. In effect, the mediums get to "choose" which questions to answer and which to leave blank (i.e., leave certain facts unmentioned), whereas the students are forced to answer every specific question. Would a medium who had not mentioned the cause of a child's death have done any better than the students if required specifically to answer the question "What caused the child's death?" Perhaps. Perhaps not. But that would have been the more appropriate comparison. In their critique of this experiment, Wiseman and O'Keeffe (2001) state the problem cleverly (but not quite accurately):

Conceptually, this is equivalent to testing archer skills by having someone fire an arrow, drawing a target around wherever it lands and calling it a bullseye, and then testing a "control" group of other archers by asking them to hit the same bullseye. (p. 29)

Schwartz conducted a follow-up experiment designed to lessen the problem of sensory cues by not permitting the sitter to respond aloud to the medium's questions. The sitter was seated behind the medium and remained silent for the first 10 minutes of the reading. Although 4 mediums and 10 sitters participated, only 1 of the sitters was able to return to the laboratory to provide quantitative ratings. She assigned a +3 rating to 77% of the statements made by the medium during the silent period and to 85% of the statements made when she answered the questions aloud. Schwartz again compared these figures with the control data obtained from the student questionnaires in the original experiment.

Neither of these experiments, of course, completely rules out sensory leakage because the sitter and the mediums are not isolated in separate rooms, a precaution that has long been routine practice in psi research (e.g., the ganzfeld telepathy experiments).

The next experiment in the series attempted to provide a better control baseline against which to compare the sitters' ratings. Three mediums and five female sitters participated. Again there was a silent period during which the medium received no feedback, and only the data from this period are presented. Each sitter assigned ratings not only to statements from her own readings but to statements from all the readings for the other sitters. The hypothesis was that the ratings would be higher for sitters' own readings than for the readings conducted with the other sitters. When sitters rated their own readings, the mean score was 40%, compared with a mean score 25% for the readings of other sitters, a significant difference, $p < .03$. (Schwartz offers some hypotheses about why the hit rates are so much lower in this experiment than in the previous experiments.)

Although this experiment provides a better chance baseline, it is still marred by the fact that the sitter knows which of the readings were hers and which of the readings were for others. This can produce several biases in favor of higher ratings for her own readings. For example, if she comes across an ambiguous or puzzling statement in one of her own readings, she will be prompted to search her memory more assiduously for confirmation (e.g., "Oh, I see now. This statement must be referring to my *other* grandmother.") She will not process statements in readings for other sitters with the same thoroughness.

A final set of two experiments used only the medium Laurie Campbell, who had participated in the earlier studies. In her own private practice, she often conducts telephone readings with her clients, and she had begun to notice that she was receiving information even before a reading got underway. So she began to do 15-minute prereading meditations in which she would write down whatever information came to her and then go over that information with the client once the phone call had begun. She followed this up with the usual reading. Schwartz adapted her technique for an experiment that controlled for sensory leakage. Three sitters would remain in their homes waiting to be called for their readings. Half an hour before a scheduled reading, Campbell would meditate and write down her impressions (defined as Phase 1, the prereading period). The sitter was then called and told that the medium would come on the line but that the phone would be muted during a 10-minute silent-reading period so that the sitter would not be able to hear the medium's reading (defined as Phase 2, the silent-reading period). At the end of the silent period, the phone was taken off mute, and Campbell introduced herself and read aloud the notes she had made prior to the phone session (i.e., during the prereading). The information from the silent-reading period was not disclosed to the sitter so that it could be preserved for later scoring. Finally, the medium completed the reading using the usual kind of back-and-forth dialogue with the sitter (Phase 3).

Unfortunately, the reporting of the results from this experiment is maddeningly incomplete, both in the text and in Appendix E, where the

published archival report (Schwartz & Russek, 2001) is reproduced. In the latter, the authors state:

A full report summarizing group statistics is in preparation. Statistically significant evidence for anomalous information retrieval was found for each of the three sitters investigated in the experiment. However, it is the uniqueness and extraordinarily evidential nature of the particular reading highlighted in this detailed report that justifies focusing on this ... research reading. (p. 336)

In other words, data from only one sitter, George Dalzell, is reported. Even in this case, however, no results are reported from the systematic rating procedure used in previous experiments. Instead, the authors present a table listing all the names mentioned by the medium during the reading, accompanied by Dalzell's ratings of their accuracy, using a scale from 0 to 3. Of the 31 names or variations, only 3 had no connection with Dalzell (i.e., he assigned them a score of 0) and 13 were absolutely accurate (assigned a score of 3). There is also a lengthy discussion of the many details revealed by the medium that Dalzell confirmed as accurate. His overall subjective estimate was that the medium was correct on 90% of her statements.

Certainly the qualitative results for this one sitter appear very impressive, even dazzling, and one can appreciate why all the participants were so excited by the outcome. But the effectiveness of the control procedure in Phases 1 and 2 during which the medium and sitter did not communicate was compromised in Phase 3, during which the medium and the sitter communicated in the usual way. It appears that the information from the earlier two phases was disclosed to Dalzell before he made his subjective accuracy estimates. The major problem, however, is that there is no systematic quantitative evaluation or statistical analysis to back up the results.

The final experiment reported in the book was designed to correct all the shortcomings of previous experiments. It was an improved replication of the experiment just described. Once again, all communication was done over the phone, with the medium (Laurie Campbell) doing the readings from Tucson, AZ, and six sitters in their own homes in different parts of the country. George Dalzell was again one of the sitters. The medium was not told who the sitters would be, and neither she nor the sitters ever heard each other's voices. Moreover, the experimenters were uninformed as to the order in which the six sitters were to be run. The tapes of Campbell's readings were transcribed, and each sitter was mailed two unmarked transcripts; one was his or her own reading and the other was a transcript of a reading done for one of the other sitters.

As with the previous experiment, however, the reporting of the results is very sketchy. Here is all we are told: "The findings were

breathhtaking. Once again it was George Dalzell's reading [that] stood out" (p. 236). We are also told that Dalzell gave accuracy ratings of 60% and 65% for the prereading information and silent period information, respectively, compared with ratings of 0% for both phases of the control reading. That's it for the results.

This experiment was subsequently reported more fully in Schwartz, Geoffrion, Shamini, Lewis, and Russek, (2003). The six sitters were instructed to read through the two transcripts, circling every item they judged to be a "dazzle shot," which was defined as "some piece of information—whatever it is TO YOU, that you experience as 'right on' or 'wow' or 'that's my family'" (p. 130). 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 attempted to designate which of the two transcripts was actually based on his or her own reading. The hypotheses were that if the medium were actually accessing information from the sitters' deceased acquaintances, then sitters should be able to identify which of the two transcripts was from their own session, and they should be able to identify significantly more "dazzle" shots, more hits, and fewer misses in their own transcript than in the control transcript.

Only four of the six sitters correctly identified their own transcripts, and there were no significant differences in the numbers of dazzle shots or hits and misses. As the authors note, "there was no apparent evidence of a reliable anomalous information retrieval effect" (p. 119). Despite this disappointing result, the pairing of medium Campbell with sitter Dalzell was once again spectacularly successful. It should be noted that Dalzell is himself a practicing medium who has written his own book about remaining in contact with his deceased romantic partner (Dalzell, 2002).

As noted at the beginning of this review, Schwartz regards the cumulative findings from his experiments as compelling evidence for the afterlife hypothesis whereas critic Hyman regards them as totally unconvincing. My own assessment falls somewhere between these two views. I believe that the experiments do reveal evidence for some form of anomalous communication that cannot be attributed to fraud, sensory leakage, the inadequacies of subjective evaluations, or flawed statistical assessments. Despite the criticisms expressed in this review, I find sufficient support in the quantitative findings to be open to the validity of the "dazzle" shots quoted in the text.

I remain unconvinced, however, that the experiments demonstrate communication with the deceased. I do not believe that the data are better explained by the afterlife hypothesis than by psi among the living participants. As noted in this review, Schwartz's explicit attempts to rule out the psi hypothesis failed. Medium Laurie Campbell was able to describe the control picture drawn by medium Suzy Smith as accurately as the pictures allegedly suggested by the deceased, and she successfully "remote viewed" the interior of SS's apartment. In the experiment with Schwartz himself, she was able to answer questions about living persons as accurately as questions

about deceased persons. And in the experiment just described, significant results were found only when the sitter himself was a medium, suggesting that psi between the medium and the sitter was the mechanism involved.

There is other suggestive evidence for psi described in the book. In a chapter titled "Is There Such a Thing as Precognition?" Schwartz describes a reading conducted during the HBO experiment by the nationally known medium John Edward in which he appears to anticipate the death of the sitter's husband, who was subsequently killed in a car accident. The sitter later confided to Schwartz that prior to the Edward reading she herself had had a dream in which her husband died in a car accident. In other words, Edward's prediction—if accepted as a genuine case of anomalous communication—could have been his own precognition or a telepathic reading of the sitter's own forebodings. In an earlier chapter, Schwartz describes another case of possible precognition involving a medium's correct prediction that the sitter would receive a gift of a stuffed animal.

In the published account of the telephone experiments, Schwartz explicitly considers the possibility that psi of one sort or another might account for the results. He suggests that it cannot. He notes that the kind of language used by the mediums implies that they were not just reporting memories and images but intentions and interpretations that reflect the information processing of active "entities." He also notes that mediums occasionally report facts unknown to the sitter. His main example comes from the final experiment in which Campbell mentioned four correct facts about Dalzell's deceased acquaintances that were unknown to Dalzell at the time of the reading.

But there is a Catch 22 here. If one grants the existence of precognition, telepathy, and remote viewing—as Schwartz does—then Campbell could have obtained the information from the future when Dalzell verified the information. Because Dalzell is also a medium, he might have been the precognitive source for the then-unknown facts, and Campbell might have picked them up from him telepathically. Moreover, some of the facts were about the physical environments of the deceased, environments that continue to exist; these could have been accessed through remote viewing. The point is that it may be more difficult than it first appears to distinguish alleged communication with deceased persons from extended forms of psi among the living. Because claims for medium-mediated communication with the deceased are more extraordinary to me than claims for psi, I will always opt for the latter when both explanations can account for the data.

REFERENCES

- DALZELL, G. (2002) *Messages: Evidence for life after death*. Charlottesville, VA: Hampton Roads Publishing.

- HYMAN, R. (2003). How *not* to test mediums: Critiquing the afterlife experiments. *Skeptical Inquirer*, 27, 20-30.
- SCHWARTZ, G.E. (2003). How *not* to review mediumship research. *Skeptical Inquirer*, 27, 58-61.
- SCHWARTZ, G.E., GEOFFRION, S., SHAMINI, J., LEWIS, S., & RUSSEK, L. (2003). Evidence of anomalous information retrieval between two research mediums: Replication in a double-blind design. *Journal of the Society for Psychical Research*, 67, 115-130.
- SCHWARTZ, G.E.R., & RUSSEK, L. (2001). Evidence of anomalous information retrieval between two research mediums: Telepathy, network memory resonance, and continuance of consciousness. *Journal of the Society for Psychical Research*, 65, 257-275.
- WISEMAN, R., & O'KEEFE, C. (2001). Accuracy and replicability of anomalous after-death communication across highly skilled mediums: A critique. *The Paranormal Review*, 19, 3-6.

DARYL J. BEM

Department of Psychology
Cornell University
Ithaca, NY 14850
d.bem@cornell.edu