

CONTRA GEORGE HANSEN'S FLAWED CRITIQUE OF THE WORK WITH B.D.

BY EDWARD F. KELLY

George Hansen has alluded on several previous occasions (e.g., 1988, 1990) to supposed weaknesses in our work with B.D. More recently (1991, 1992), however, this has become for the first time a full-scale attack, one which Hansen apparently believes to be highly successful and which he therefore uses as a stepping-stone toward a larger thesis concerning the proper role of magic and magicians in psi research.

I will confine my response largely to Hansen's discussion of the formal publications describing our research with B.D. In brief, although I have qualified sympathy for the larger thesis, I have virtually none for his treatment of our experimental work or for his abuse of others for their confidence in it.

Let me begin by giving a correct account of the overall flow of the work. There were two periods of intense activity. The first occurred in February and March of 1972 during B.D.'s first visit to Durham and was chiefly directed toward (a) determining B.D.'s capacity for psi performance under controlled conditions, and (b) identifying potentially productive directions for systematic research. The main experiments—the single-card clairvoyance and shuffles experiments—took place between October 1972 and May 1973 during B.D.'s leave of absence from Yale Law School. This later research was supported by the Hodgson Fund, following B.D.'s appearance at Harvard at the session described in Persi Diaconis's 1978 *Science* paper, a subject to which I will return.

The work carried out during the first period (Kelly & Kanthamani, 1972) consisted not of *experiments* in the normal sense but of *tests* conducted with a wide variety of standard devices and procedures then in use at the FRNM. Foremost among these early efforts was the work with Helmut Schmidt's four-button machine. Hansen

This paper is a revised version of remarks originally prepared for oral presentation at the 1991 PA Convention by invitation of the Program Chair, but withdrawn when Hansen did not appear to deliver his own paper.

I wish to express my gratitude to Irvin L. Child for his help and constructive criticism throughout this project and to John Beloff, Gertrude Schmeidler, and John Palmer for reading earlier drafts of this paper and providing their comments.

dismisses all these results in a single short paragraph, primarily by reference to the Radin and Nelson (1987) meta-analysis of REG work, which he says "assigned these studies some of the lowest possible quality ratings." This seems to me an entirely inappropriate use of Radin and Nelson's rating scheme, which is concerned primarily with assurances of REG quality needed to support reliable measurement of small deviations from mean chance expectation in large numbers of trials (the normal situation even in experiments with subjects preselected for ability in the pertinent task). By contrast, consider B.D.'s first session in the FRNM library, with J. B. Rhine and Helmut Schmidt observing along with Kanthamani, myself, and other members of the FRNM staff. During the course of this get-acquainted meeting, in which he mainly chatted with us about himself and his views of psi, B.D. intermittently made responses on the four-button machine. Over the course of the approximately hour-long session he accumulated a total of 180 hits in 508 trials, for a scoring rate of 35.4%. (These figures, incidentally, were corroborated by the independent duplicate counters sealed inside the device.) The probability for this one outcome is under 10^{-6} . Had its occurrence depended on chance alone, the meeting might well have gone on for something on the order of a million hours (since the expected waiting time for an event of probability p is $1/p$). Even stronger performances were observed less formally, and the visit ended with a formal series of eight sessions, recorded both manually and on paper tape, in which B.D. progressively raised his scoring rate from 27% to almost 31% ($CR > 6$, $p < 10^{-9}$).

These very unusual scores, moreover, cannot plausibly be attributed to hypothetical failures of the test device. The physical device in question is the quantum-mechanical REG originally developed and tested at Boeing by Helmut Schmidt and brought by Schmidt to the FRNM for use in his continuing research. In addition to the various mechanical and electronic safeguards built into it, the device had been subjected to an unusually thorough regime of randomness testing specifically directed to the forms of potential failure that could be anticipated in light of its electronic design. Schmidt's original reports (1969, 1970) should be consulted for details, but control tests involving 4.7×10^6 numbers were generated on 100 different days over an 18-month period, interspersed with on-going experimental sessions and using a generation rate close to B.D.'s own routine response rate of about one per second. The results were analyzed both overall and in blocks of various sizes, including blocks approximating the size of B.D.'s tape-recorded sessions. Additional

control tests were carried out using the tape-recorded guess sequences of successful subjects to trigger the REG at a variety of input rates. In none of these tests was there any sign of significant departure from ideal randomness. Additional, more routine randomness checks conducted both before and after B.D.'s visit lend no support whatsoever to any casual speculation that local or global failures could have occurred on a scale sufficient to explain the extreme scores produced by B.D. I should also point out that large numbers of other persons, including myself, were trying hard with the same device, and in the same time frame, with very little success.

In short, I categorically reject Hansen's offhand dismissal of the REG results with B.D. Furthermore, Helmut Schmidt himself, who was present throughout B.D.'s first visit, is equally uncompromising in the view that this REG work by itself conclusively established B.D.'s capacity for controlled psi performance (personal communication, August 1, 1991).

I have dwelled on the REG work at length both for its own sake and because it formed for us a crucial part of the context in which the subsequent card experiments evolved (Kanthamani & Kelly, 1974a, 1974b, 1975). For us the goal of these later experiments was not to demonstrate that B.D. "had psi," but to learn something significant about its modes of operation. In that context, we deliberately sought to establish experimental conditions that would enable us to elicit strong performance from B.D. even if those conditions were less than ideal, *provided*, however, that they did not compromise in any fundamental way the integrity of the experiments, even on the assumption that the subject would be inclined and able to cheat if given the opportunity. This is the spirit in which the card experiments were designed and conducted. The only point at which we bent these rules even further was in the later series of the shuffle experiments, which certainly do not stand on their own in terms of the quality of conditions. However, in the published reports we ourselves clearly identified these weaker series, explained why we permitted them in light of B.D.'s previous performances, and segregated their results in analysis and discussion.

Hansen construes not only these "special" modifications but every other aspect of protocol originated by B.D. as presumptive evidence that B.D. obtained thereby an opportunity to cheat, and that he did in fact cheat. It is crucial to recognize, however, (a) that the alleged "flaws" are very abstract in character, and (b) that Hansen provides no direct, positive evidence that they actually occurred in a form that enabled cheating to occur. In fact, there are strong

counterarguments against all of Hansen's allegations, and especially against those involving the single-card clairvoyance series, which we too regard as the most significant part of the later work. Let me therefore comment briefly on each of the four major "flaws" Hansen claims to have identified.

1. *Commotion.* Hansen quotes a statement we made in our first paper regarding the need for sometimes heated arguments to convince B.D. to work in an experimental setting (Kelly & Kanthamani, 1972, p. 88) and uses it repeatedly to paint a picture of experimental sessions routinely characterized by commotion and distractions that provided B.D. with cover for his "moves." Such accusations are grossly inaccurate, however: Even during the initial visit, the argumentation—entirely between myself and B.D., by the way—took place *outside* the context of the formal testing; for example, during long walks around Duke's campus, trips between the lab and my house (where B.D. was staying), and late-night discussions of what we as experimenters were trying to do, and why. By the time of the later card experiments, B.D. had become much more tolerant of experimental requirements (although he still grumbled about them, and clearly found the work stressful), and these sessions were routinely professional and quiet.

2. *Confederates.* The primary potential exception to these statements concerns a subset of sessions in which additional persons were present. Hansen speculates that this could have permitted confederates in the room to glimpse the target cards and signal them to B.D. This suggestion also is without merit. Even had confederates been present, I very strongly doubt that they would have been able to "glimpse" anything (as discussed in more detail below), and there are very few candidates in sight. According to our records only two sessions introduced true "visitors" in the sense required by Hansen's criticism. One of these involved as its single visitor a law-school classmate of B.D., who certainly must be regarded as a potential confederate, and the scoring for this session was good although far from the best in the series. The other involved a group of students from Tennessee, and although conditions were optimal in this session for B.D. to take advantage of commotion and confederates, it actually produced some of the worst scores of the entire experiment (0 exact, 1 number, and 3 suit hits in 15 trials).

3. *Recording errors.* Repeating a suggestion previously made by Akers (1986), Hansen speculates that the recording procedure we adopted might have permitted B.D. to glimpse the target before he made his response, thus providing another example of how aspects

of protocol suggested by the subject enabled him to cheat. It should be pointed out first that this suggestion flatly contradicts the clear statement, repeatedly made in the published reports, that the experimenter *first* recorded the call and *then* displayed and recorded the target. This was easily accomplished because of the unusually slow rate at which trials were generated (typically on the order of one per minute). That an experienced experimenter such as Kanthamani would permit numerous breaches of this essential feature of protocol is simply not credible in the absence of positive evidence. In this regard, Akers's comments are not "especially weighty" (as Hansen contends) because (a) he did not witness the main experiments, but only a brief series carried out subsequent to my departure from the FRNM, and (b) even in that context he did not actually observe the hypothesized breach of protocol. Furthermore, this "flaw" in principle does not apply to the batches of no-feedback trials interspersed through the series, for which the recording of the targets was carried out separately after the recording of the calls. Nevertheless, the scoring for these (179) trials was actually substantially *higher* than it was for the normal feedback (289) trials of the same runs (4.36 vs. 2.34 exact hits and 9.0 vs. 7.02 number hits per 52 trials, with mean chance expectation being 1 and 4, respectively).

4. *Visual leakage.* Hansen's most serious allegation, in my estimation, is that B.D. might have used or arranged reflective surfaces within the experimental room to obtain occasional glimpses of the target cards as they were being selected and presented by the experimenter. Hansen intimates that we did not consider such possibilities, and he interprets our finding of visual-like confusion patterns in the clairvoyance data (Kelly, Kanthamani, Child, & Young, 1975) as evidence that visual leakage did in fact occur. This finding, however, is not an incidental result but the *main* result of the single-card clairvoyance study, the principal goal of which was to compare the systematic errors B.D. made under visual versus ESP conditions. This central comparison is obviously devoid of value unless the conditions of the ESP trials precluded visual access to the targets. To this end, we adopted a set of conditions specifically developed for this purpose by Irvin Child during his own earlier (and only marginally successful) work with B.D. To recapitulate briefly, successive targets were drawn from a pool of 10 decks of ordinary playing cards. The cards were thoroughly shuffled and arranged sideways, with their backs toward the experimenter, in a cardboard box kept in the bottom drawer of a large solid-backed office desk. The experimenter "randomly" selected a card and, with its face down, par-

allel to the floor, inserted it into an oversized black folder. I emphasize that the protocol required this target selection and masking to take place entirely within the bottom drawer, below the height of its walls, and was intended to assure that neither the subject nor the experimenter could have visual access to the target. The folder was then held up and displayed to B.D. with the back of the card facing him inside it.

I assert categorically, on the basis of our examination of this possibility prior to initiating the experiments, that under these conditions there existed *no* ready-made optical paths enabling B.D. or anyone else in the room to glimpse a target by way of windows, doorknobs, spectacles, or other reflective surfaces routinely present.¹ I also find it extremely implausible that B.D. could successfully have introduced, and used on numerous occasions without detection, an additional optical path or paths of his own construction. Such a path would necessarily have been complex and would have involved an outer wall or the floor of the drawer holding the cards. As experimenters and observers, we were in and around that drawer virtually every day, and at unpredictable times, and we never detected any trace of tampering with either the drawer, the desk, or the cards themselves. I should also mention here that we had told B.D. on a number of occasions (particularly during his first visit) that if he were ever caught cheating we would immediately terminate the experiments and renounce all previous work with him. In sum, although this alleged "flaw" is more open-ended than the others and we cannot claim to have dispatched it as conclusively, I believe that any fair-minded observer familiar with the experimental procedures and the physical setting would conclude with us that the visual-access hypothesis is not tenable, particularly in the generic form that Hansen advances unaccompanied by any specific proposal as to how such access could have been obtained. Indeed, to my mind the leakage hypothesis that we ourselves originally suggested and rejected—that is, leakage arising from breaches of the card-handling protocol (Kelly et al., 1975)—is the *least* implausible of the various non-psi hypotheses offered to date in explanation of the ESP confusion patterns we observed.

I will say little about the shuffles data beyond what I have already said. The failure of the visual-like confusion pattern to appear in these data, which Hansen interprets simply as evidence that the

¹ See also the more detailed description of the physical setting by H. Kanthamani presented in an accompanying article.

cheating in this series took a different form, has a good alternative interpretation: It was a different task, and B.D.'s achievement of high scores with small numbers of shuffles provides a statistical argument for construing the psi effect in this series as a PK effect rather than an ESP effect (see Kanthamani & Kelly, 1975). Furthermore, if the visual-like pattern *had* appeared, I am sure Hansen would have been equally quick to interpret it as evidence that B.D. cheated by glimpsing cards through the holes in the box or by momentarily exposing their edges. I would also like to point out in this connection that Hansen conspicuously neglects to mention two runs from this series that were specifically immune to the general form of cheating he suggests (inasmuch as B.D. did not touch the cards after his shuffling was completed). Both of these runs yielded even higher scores than the series at large, with 5 and 7 exact hits representing independent Poisson probabilities of .003 and .00007, respectively.

This completes the main outline of my responses to the substance of Hansen's critique; but before concluding, I also want to comment on what I perceive as an underlying double standard in terms of Hansen's willingness to accept without apparent question, in support of his own views, the unqualified defamatory statements regarding B.D. that have been issued by magicians such as Randi, Gardner, and especially Persi Diaconis. I find it particularly galling in this regard that Hansen chastises me for not inviting Diaconis to participate in the formal studies. Let me immediately set the record straight on that: I had never heard of Persi Diaconis until he published his paper in *Science* in 1978, six years after the session at Harvard and five years after completion of the card experiments. Had he approached me at the Harvard session and offered his services, I would have accepted on the spot. However, he apparently felt no obligation even to introduce himself, let alone to inform me regarding his suspicions.

The background of Diaconis's involvement in the Harvard session may also be of interest, albeit less certain historically. I had initially invited his statistical mentor, Fred Mosteller, to attend the meeting, but to my surprise Mosteller heatedly refused, on grounds that in his opinion Edgar J. Coover had already conclusively demonstrated in the 1920s that ESP research is all snare and delusion! I suspect that Mosteller then dispatched Diaconis to that session as his agent, and with deliberate secrecy, for the sole and specific purpose of "exposing" Bill Delmore. I further suspect that Mosteller, a former president of the AAAS, was instrumental both in *Science's*

publication of Diaconis's 1978 paper and in their refusal to follow the editorial practices stated on their own masthead in regard to the detailed reply that I submitted *as a paper* the following week (Kelly, 1979).²

Ironically, Diaconis himself characterized our published papers as describing experimental conditions "beyond reproach," but dismissed their results on grounds that the *actual* conditions must have differed radically (in some unspecified way) from those we described. The basis for this astonishing suggestion rests on a deliberate and untruthful characterization of the Harvard session, which Diaconis knew was completely informal, as not just one but a whole series of "experiments" in the normal sense. To my knowledge Diaconis has never directly addressed the details of our published experimental reports; indeed, I have some doubt, personally, that he had even *read* them at the time he wrote his *Science* paper, which in my opinion falls far beneath the customary standards of that journal (Kelly, 1979). Nevertheless, people such as James Randi and Martin Gardner—and no doubt numerous others by now—refer to his *Science* paper with unqualified approval. Can anyone seriously imagine these people to be open-minded students of the work with B.D.? In effect, they will use every means available, including unconstrained appeals to the powers of magicians, to dismiss any experimental outcomes they do not like. Involvement of magicians in work with special subjects can perhaps offer *some* protection against this sort of thing, and to that extent is desirable; but it is simply naïve to imagine that it can provide immunity. Instead, it simply raises new issues about the relative credibility of the various magicians who might line up on either side.

To conclude, I submit that we did in fact "reasonably exclude" cheating in our work with B.D. and that Hansen's critique has little or no real substance. In fairness, however, I must also acknowledge that some of Hansen's misperceptions were certainly encouraged, or at least not specifically *discouraged*, by what appears to me in retrospect to be a definite failure on our part to report sufficient detail on a number of significant points. I can and do thank him, therefore, for enabling us to rectify these deficiencies while we are still in condition to do so. I also think there is an underlying issue here, the open discussion of which could represent another positive outcome of this exchange. We deliberately chose in our write-ups of the

² The only response I ever received from *Science* was a phone call several months later from the "Letters" editor, who wanted to discuss which one paragraph they should select from my nine-page rebuttal (Kelly, 1979).

work with B.D. to avoid dwelling in a paranoid way on our defenses against possible attempts at cheating on the part of the subject. In hindsight I am inclined to think we went somewhat too far in that direction, to the ultimate detriment, perhaps, of both B.D. and ourselves. However, I would still argue strongly for a middle ground between our approach and that exemplified by Honorton's recent report of his work with Malcolm Bessent (Honorton, 1987), in which the central purpose and outcome of the experiment practically disappear beneath the welter of precautionary details. There is certainly room for expression of personal preference in these matters, but it might also be appropriate for the Parapsychological Association to try to establish guidelines for the conduct and reporting of future research involving special subjects. Collaboration with professional magicians could be useful, I think, provided that the participating individuals are not only technically qualified, but also emotionally and intellectually capable of entertaining the possibility that genuine psi phenomena do occur.

REFERENCES

- AKERS, C. (1986). Has parapsychology found its basic experiments? [Review of *The basic experiments in parapsychology*, K. R. Rao (Ed.)]. *Contemporary Psychology*, **31**, 180–181.
- DIACONIS, P. (1978). Statistical problems in ESP research. *Science*, **201**, 131–136.
- HANSEN, G. P. (1988). Risks of deception by subjects. *Proceedings of presented papers: The Parapsychological Association 31st annual convention* (pp. 163–179).
- HANSEN, G. P. (1990). Deception by subjects in psi research. *Journal of the American Society for Psychical Research*, **84**, 25–80.
- HANSEN, G. P. (1991). The research with B.D. and the legacy of magical ignorance. *Proceedings of presented papers: The Parapsychological Association 34th annual convention* (pp. 172–188).
- HANSEN, G. P. (1992). The research with B.D. and the legacy of magical ignorance. *Journal of Parapsychology*, **56**, pp. 307–333.
- HONORTON, C. (1987). Precognition and real-time ESP performance in a computer task with an exceptional subject. *Journal of Parapsychology*, **51**, 291–320.
- KANTHAMANI, H., & KELLY, E. F. (1974a). Awareness of success in an exceptional subject. *Journal of Parapsychology*, **38**, 355–382.
- KANTHAMANI, H., & KELLY, E. F. (1974b). Card experiments with a special subject: I. Single-card clairvoyance. *Journal of Parapsychology*, **38**, 16–26.
- KANTHAMANI, H., & KELLY, E. F. (1975). Card experiments with a special subject: II. The shuffle method. *Journal of Parapsychology*, **39**, 206–221.

- KELLY, E. F. (1979). Reply to Persi Diaconis. *Zetetic Scholar*, No. 5, pp. 20-28.
- KELLY, E. F., & KANTHAMANI, B. K. (1972). A subject's efforts toward voluntary control. *Journal of Parapsychology*, 36, 185-197.
- KELLY, E. F., KANTHAMANI, H., CHILD, I. L., & YOUNG, F. W. (1975). On the relation between visual and ESP confusion structures in an exceptional ESP subject. *Journal of the American Society for Psychical Research*, 69, 1-31.
- RADIN, D. I., & NELSON, R. D. (1987). *Replication in random event generator experiments: A meta-analysis and quality assessment* (Human Information Processing Group Technical Report 87001). Princeton, NJ: Princeton University.
- SCHMIDT, H. (1969). *Anomalous prediction of quantum processes by some human subjects*. (Boeing Scientific Research Laboratories Document D1-82-0821). Seattle, WA.
- SCHMIDT, H. (1970). Quantum-mechanical random-number generator. *Journal of Applied Physics*, 41, 462-468.

*c/o Institute for Parapsychology
P. O. Box 6847, College Station
Durham, NC 27708*