misses estimated; perhaps Cohen's h greatly underestimates effect size when very low probability events (less than 1 in 50 for heart attack in the placebo condition and less than 1 in a 100 for aspirin) are involved. I'm not a statistician and thus don't know if there is a relevant literature on this point.

The above objections should not detract from the overall value of the Utts survey. The findings she reports will need to be replicated; but even as is, they provide a challenge to some of the cherished arguments of counteradvocates, yet also challenge serious researchers to use these findings effectively as guidelines for future studies.

Comment

Frederick Mosteller

Dr. Utts's discussion stimulates me to offer some comments that bear on her topic but do not, in the main, fall into an agree-disagree mode. My references refer to her bibliography.

Let me recommend J. Edgar Coover's work to statisticians who would like to read about a pretty sequence of experiments developed and executed well before Fisher's book on experimental design appeared. Most of the standard kinds of ESP experiments (though not the ganzfeld) are carried out and reported in this 1917 book. Coover even began looking into the amount of information contained in cues such as whispers. He also worked at exposing mediums. I found the book most impressive. As Utts says in her article, the question of significance level was a puzzling one, and one we still cannot solve even though some fields seem to have standardized on 0.05.

When Feller's comments on Stuart and Greenwood's sampling experiments came out in the first edition of his book, I was surprised. Feller devotes a problem to the results of generating 25 symbols from the set a, b, c, d and e (page 45, first edition) using random numbers with 0 and 1 corresponding to a, 2 and 3 to b, etc. He asks the student to find out how often the 25 produce 5 of each symbol. He asks the student to check the results using random number tables. The answer seems to be about 1 chance in 500. In a footnote Feller then says "They [random numbers] are occasionally extraordinarily obliging: c.f. J. A. Greenwood and E. E. Stuart, Review of Dr. Feller's Critique, *Journal of Parapsychology*, vol. 4 (1940), pp. 298–319, in particular p. 306." The 25 symbols of 5 kinds, 5 of each, correspond to the cards in a parapsychology deck.

The point of page 306 is that Greenwood and Stuart on that page claim to have generated two random orders of such a deck using Tippett's table of random numbers. Apparently Feller thought that it would have taken them a long time to do it. If one assumes that Feller's way of generating a random shuffle is required, then it would indeed be unreasonable to suppose that the experiments could be carried out quickly. I wondered then whether Feller thought this was the only way to produce a random order to such a deck of cards. If you happen to know how to shuffle a deck efficiently using random numbers, it is hard to believe that others do not know. I decided to test it out and so I proposed to a class of 90 people in mathematical statistics that we find a way of using random numbers to shuffle a deck of cards. Although they were familiar with random numbers, they could not come up with a way of doing it, nor did anyone after class come in with a workable idea though several students made proposals. I concluded that inventing such a shuffling technique was a hard problem and that maybe Feller just did not know how at the time of writing the footnote. My face-to-face attempts to verify this failed because his response was evasive. I also recall Feller speaking at a scientific meeting where someone had complained about mistakes in published papers. He said essentially that we won't have any literature if mistakes are disallowed and further claimed that he always had mistakes in his own papers, hard as he tried to avoid them. It was fun to hear him speak.

Although I find Utts's discussion of replication engaging as a problem in human perception, I do always feel that people should not be expected to carry out difficult mathematical exercises in their head, off the cuff, without computers, textbooks or advisors. The kind of problem treated requires careful formulation and then careful analysis. Even

Frederick Mosteller is Roger I. Lee Professor of Mathematical Statistics, Emeritus, at Harvard University and Director of the Technology Assessment Group in the Harvard School of Public Health. His mailing address is Department of Statistics, Harvard University, Science Center, 1 Oxford Street, Cambridge, Massachusetts 02138.