Figure 14.1. TESTS OF SIGNIFICANCE: Investigating Extrasensory Perception

EM9538: The Globe and Mail, December 8, 1995, page A12

To skeptics, seeing isn't believing

Were it not in such an unusual domain, extrasensory perception would no longer be questioned by science as a real phenomenon, says a researcher who recently presented statistical evidence

BY STEPHEN STRAUSS
Science Reporter
Toronto

HOW do you establish statistical proof for something as contentious as extrasensory perception? With great, great difficulty and with many doubters in your wake.

Last week, Jessica Utts, a professor of statistics at the University of California at Davis, unveiled her analysis of 20 years of ESP experiments conducted for the Central Intelligence Agency and other U.S. intelligence agencies.

There was, she said, a small to medium-sized ESP effect that was statistically robust. "Were it not in such an unusual domain, [ESP] would no longer be questioned by science as a real phenomenon."

Other statisticians responded to her assertion with an interesting mix of agreement and disagreement. University of Oregon psychologist Ray Hyman, a longtime parapsychology critic, said the results "were free of the methodological weaknesses that plagued the early..... research" and were "too large and too consistent to be dismissed as statistical flukes."

For Prof. Utts to persuade her colleagues even so far was a considerable accomplishment in a field littered with both charlatans and repeated methodological blunders.

In an attempt to get away from the levitations and seances of psychic mediums, early parapsychologists tried to use card prediction as a test of precognition. Then it was pointed out to them that a random shuffle was almost impossible to achieve and that subjects with good memories could simply remember the order of the previous deal, not to mention the hints given by tiny imperfections in the backs of cards.

A problem less easily confronted is known as the "file-drawer effect." Researchers generally published only their positive results and consigned the rest of their data to their files. Statistically speaking, this amounted to reporting only the flukes, the improbable results that do sometimes crop up, just as 7 may, very infrequently, turn up five times in a row at the craps table.

Then there was the repeatability problem. Not even the best of the supposedly psychic subjects seemed able to turn on their skills at will, therefore you could never repeat the same experiment and know you were getting the same result for the same reason.

Responding to these and other criticisms, parapsychologists at research institutes SRI International and Science Applications International Corp., funded by the U.S. intelligence agencies, went to great lengths to produce statistically unassailable data.

A person was asked to draw or describe a place, object or picture that a sender was mentally transmitting to him or her.

Sometimes, the images to be transmitted were chosen at random. For example, a sender might go to a windmill farm and send back an image of that.

Other times, the process was more constricted and used only pictures taken from *National Geographic* magazine. A person would try to sketch a picture, selected by computer, which he or she had not seen. The picture was then packaged with three or four others. Careful attention was given to ensuring that each packet contained images very different from each other. For example, no two *National Geographic* pictures would show animals.

Then an independent judge, who did not know which image had been chosen, would compare the receiver's sketch or description with the images and decide which was the most likely match.

If there were four pictures in each packet, chance would say the judge had a one-in-four likelihood of picking the correct one. Looking at large numbers of such trials conducted by Science Applications, both Prof. Utts and Prof. Hyman concluded that people were constantly scoring above chance.

How much above chance is difficult to quantify because a number of test designs were chosen. In the one-in-four example, the rate of success was closer to one in three.

Some people were on occasion very good: correct as much as 60 to 70 per cent of the time. There was no evidence that you could improve people's psychic abilities. Distance between sender and receiver made no difference. About 1 per cent of the people tested seemed to have some psychic abilities. Electromagnetic shielding had no effect.

Yet, said Stephen Fienberg, a Canadian statistician who teaches at Carnegie Mellon University in Pittsburgh, "We still have a way to go before we can conclude that parapsychologists have convincingly demonstrated the existence of psychic phenomena."

Part of the reason for the continued disbelief is that purely statistical proof stands the usual order of science on its head. Generally in science, a phenomenon is observed – the sun shining, a wiggle in the orbit of a planet – and then scientists try to understand it. In psychic research, because results can vary from time to time even with the same people, all you have is a statistical result that is above average.

"A .300 hitter in baseball doesn't get a hit every third time he bats, but when he gets a hit, you know it is a hit" is how Prof. Hyman characterizes the difficulties.

In psychic research, you can't separate a true hit from a chance occurrence that looks like a hit.

Prof. Utts has tried to counter this criticism by pointing out that in large-scale epidemiological studies – the effect of a small dose of acetylsalicylic acid on heart attacks, for example – the protective effect was so relatively small that it could be seen only when tens of thousands of people participated in an experiment.

Her critics point out that for that finding to be convincing, people who took ASA had to be compared with people who didn't. This allowed researchers to arrive at what is called the "null hypothesis." It wasn't just the taking of a pill that brought the heart-attack rate down; it was something in the ASA.

"Psychic phenomena must be one of the most difficult things in the world to establish a null hypothesis for," said Wayne Oldford, a statistician at the University of Waterloo in Ontario.

You lack a control group of people who you are certain are receiving no images and whose guesses must entirely represent chance.

Beyond all this is the lack of a theoretical framework to explain the physics of the act. How is an image sent; where is it received in the body and why is its success rate so

(*continued overleaf*)