

Control group study, double-blind and random tests

A control group study uses a control group to compare to an experimental group in a test of a causal hypothesis. The control and experimental groups must be identical in all relevant ways except for the introduction of a suspected causal agent into the experimental group. If the suspected causal agent is actually a causal factor of some event, then logic dictates that that event should manifest itself more significantly in the experimental than in the control group. For example, if 'C' causes 'E', when we introduce 'C' into the experimental group but not into the control group, we should find 'E' occurring in the experimental group at a significantly greater rate than in the control group. Significance is measured by relation to chance: if an event is not likely due to chance, then its occurrence is significant.

A double-blind test is a control group test where neither the evaluator nor the subject knows which items are controls. A random test is one which randomly assigns items to the control or experimental groups.

The purpose of controls, double-blind and random testing is to reduce error, self-deception and bias. An example should clarify the necessity of these safeguards.

The DKL LifeGuard Model 2, from DielectroKinetic Laboratories, can detect a living human being by receiving a signal from the heartbeat at distances of up to 20 meters through any material, according to its manufacturers. Sandia Labs tested the device using a double-blind, random method of testing. Sandia is a national security laboratory operated for the U.S. Department of Energy by the Sandia Corporation, a Lockheed Martin Co. The causal hypothesis they tested could be worded as follows: the human heartbeat causes a directional signal to activate in the LifeGuard, thereby allowing the user of the LifeGuard to find a hidden human being (the target) up to 20 meters away, regardless of what objects might be between the LifeGuard and the target.

The testing procedure was quite simple: five large plastic packing crates were set up in a line at 30-foot intervals and the test operator, using the DKL LifeGuard Model 2, tried to detect in which of the five crates a human being was hiding. Whether a crate would be empty or contain a person for each trial was determined by random assignment. This is to avoid using a pattern which might be detected by the subject. Their tests showed that the device performed no better than expected from random chance. The test operator was a DKL representative. The only time the test operator did well in detecting his targets was when he had prior knowledge of the target's location. The LifeGuard was successful ten out of ten times when the operator knew where the target was. It may seem ludicrous to test the device by telling the operator where the objects are, but it establishes a baseline and affirms that the device is working. Only when the operator agrees that his device is working should the test proceed to the second stage, the double-blind test. For, the operator will not be as likely to come up with an ad hoc hypothesis to explain away his failure in a double-blind test if he has agreed beforehand that the device is working properly.

If the device could perform as claimed, the operator should have received no signals from the empty crates and signals from each of the crates with a person within. In the main test of the LifeGuard, when neither the test operator nor the investigator keeping track of the operator's results knew which of five possible locations contained the target, the operator performed poorly (six out of 25) and took about four times longer than when the operator knew the target's location. If human heartbeats cause the device to activate, one would expect a significantly better performance than 6 of 25, which is what would be expected by chance.

The different performances--10 correct out of 10 tries versus 6 correct out of 25 tries--vividly illustrates the need for keeping the subject blind to the controls: it is needed to eliminate self-deception and subjective validation. The evaluator is kept blind to the controls to prevent him or her from subtly tipping off the subject, either knowingly or unknowingly. If the evaluator knew which crates were empty and which had persons, he or she might give a visual signal to the subject by looking only at the crates with persons. To eliminate the possibility of cheating or evaluator bias, the evaluator is kept in the dark regarding the controls.

The lack of testing under controlled conditions explains why many psychics, graphologists, astrologers, dowzers, paranormal therapists, etc., believe in their abilities. To test a dowser it is not enough to have the dowser and his friends tell you that it works by pointing out all the wells that have been dug on the dowser's advice. One should perform a random, double-blind test, such as the one done by Ray Hyman with an experienced dowser on the PBS program *Frontiers of Science* (Nov. 19, 1997). The dowser claimed he could find buried metal objects, as well as water. He agreed to a test that involved randomly selecting numbers which corresponded to buckets placed upside down in a field. The numbers determined which buckets a metal object would be placed under. The one doing the placing of the objects was not the same person who went around with the dowser as he tried to find the objects. The exact odds of finding a metal object by chance could be calculated. For example, if there are 100 buckets and 10 of them have a metal object, then getting 10% correct would be predicted by chance. That is, over a large number of attempts, getting about 10% correct would be

expected of anyone, with or without a dowsing rod. On the other hand, if someone consistently got 80% or 90% correct, and we were sure he or she was not cheating, that would confirm the dowser's powers.

The dowser walked up and down the lines of buckets with his rod but said he couldn't get any strong readings. When he selected a bucket he qualified his selection with something to the effect that he didn't think he'd be right. He was right about never being right! He didn't find a single metal object despite several attempts. His performance is typical of dowsers tested under controlled conditions. His response was also typical: he was genuinely surprised. Like most of us, the dowser is not aware of the many factors that can hinder us from doing a proper evaluation of events: self-deception, wishful thinking, suggestion, unconscious bias, selective thinking, subjective validation, communal reinforcement, etc.

Many control group studies use a placebo in control groups to keep the subjects in the dark as to whether they are being given the causal agent that is being tested. For example, both the control and experimental groups will be given identical looking pills in a study testing the effectiveness of a new drug. Only one pill will contain the agent being tested; the other pill will be a placebo. In a double-blind study, the evaluator of the results would not know which subjects got the placebo until his or her evaluation of observed results was completed. This is to avoid evaluator bias from influencing observations and measurements.

See related entries on ad hoc hypotheses, cold reading, communal reinforcement, confirmation bias, the experimenter effect, Occam's razor, the placebo effect, the post hoc fallacy, selective thinking, self-deception, subjective validation, testimonials, James Van Praagh and wishful thinking.

further reading

- Therapeutic Touch Study Data
- Mass Media Funk

Giere, Ronald, *Understanding Scientific Reasoning*, 4th ed, (New York, Holt Rinehart, Winston: 1998).

Kourany, Janet A., *Scientific Knowledge: Basic Issues in the Philosophy of Science*, 2nd edition (Belmont: Wadsworth Publishing Co., 1998).

Sagan, Carl. *The Demon-Haunted World: Science as a Candle in the Dark* (New York:Random House, 1995).

___x1f1@1È1