CHAPTER 3 CONTROL IN EXPERIMENTATION

3 Control In Experimentation 2

3.1 A Life Devoted To Empiricism 2

3.2 A Bold Experiment 3

3.3 Mill's Methods: A Definition Of Control 5

3.3.1 Two Steps of Research 6

3.3.2 Method of Difference 8

3.3.3 Method of Concomitant Variation 12

3.3.4 Method of Agreement 17

3.4 The Language of Variables 18

3.4.1 Independent and Dependent Variables 19

3.4.2 Quantitative and Qualitative Variables 20

3.5 Limitations Of Mill's Experimental Method 21

3.5.1 Internal Validity 21

3.5.2 External Validity 23

3.5.3 Construct Validity 24

3.6 Applying The Methods. 25

3.6.1 Within-Subjects Design 26

3.6.2 Between-Subjects Design 29

3.7 Approaches To Dealing with Uncontrolled Variables 31

3.7.1 The Within-Subjects Design with Virtually Total Control 31

3.7.2 Statistical Control: Measure the Uncontrolled Variables and Remove Their Effects 32

3.7.3 Randomized Experiments 32

3.8 Key Terms 33

3.9 Key People 34

3.10 Review Questions 35

3 CONTROL IN EXPERIMENTATION

The only thing that I believe I am really fit for, is the investigation of abstract truth, & the more abstract the better. If there is any science which I am capable of promoting, I think it is the science of science itself, the science of investigation—of method. John Stuart Mill

The studies by Franklin's commission that falsified Mesmer's theory of animal magnetism were models of clarity of thought. Presented with a bewildering set of observations and theoretical claims, the commissioners were able to cut to the heart of the matter and develop a critical test of Mesmer's ideas. Using simple techniques to *control for rival hypotheses,* they demonstrated that Mesmer's magnetic treatment on its own had no effect and that the effects attributed to it were due to suggestion; and they conducted these elegant and compelling experiments half a century before formal models for achieving control over rival hypotheses in research were even developed.

The job of formalizing principles for scientific data collection was done by John Stuart Mill. In his book *A System of Logic* (1843), Mill presented a set of abstract rules, "analogous to the rules of the syllogism," for researchers to use in reaching valid conclusions about cause and effect. If scientists follow these rules, Mill believed, their research yields conclusive results; otherwise, it does not. The rules Mill specified still set the standard of excellence in evaluating experiments and observational studies.

This chapter focuses on these rules, now known as *Mill's methods,* and illustrates how they are applied in psychological research. As we will see, modern research methods retain the logic of Mill's methods but introduce new techniques to overcome limitations in applying them.

3.1 A LIFE DEVOTED TO EMPIRICISM

John Stuart Mill described himself as a student of "the science of science itself, the science of investigation—of method" (Mill, in

Robson, 1973, p. xlix). With his work on scientific methods, Mill hoped to undermine the influence of intuitive philosophy, which held "that truths external to the mind may be known by intuition or consciousness, independently of observation and experience" (Stillinger, 1969, p. 134). He wanted to strengthen the case for empiricism, the only approach he found acceptable for verifying laws of nature.

John Stuart Mill, philosopher of science

Mill planned to identify and describe all possible procedures for discovering and verifying scientific truths by studying the methods used by great scientists like Sir Isaac Newton, Johann Kepler, and Pierre Laplace. He thought that presenting scientists with formal methods for doing research would be a marvelous accomplishment. Not only would the availability of the methods lead to faster progress in science (future scientists would have only to select the right procedure from an exhaustive list of possibilities), but the methods would further the aims of utilitarianism, the philosophy that had shaped the entire course of Mill's life. As we shall see, although John Stuart Mill was not a scientist himself, he had been intimately involved with a program of research from his earliest years.

3.2 A Bold Experiment

In the early 1800s, James Mill, John Stuart's father, a historian and writer, and Jeremy Bentham, a jurist, invented a new moral theory, which Bentham called utilitarianism. It was aimed at improving the lot of humanity through innovative educational and economic programs and new laws extending the vote to working men and women. In true scientific spirit, utilitarians held that it was not enough to judge programs on their intent, or by a rational analysis of their content, or by citing the opinions of authorities. Rather, they believed that the value of a program could be determined only by appealing to the evidence— by changing the circumstances of people's lives and observing the consequences. The best programs and laws would be those with the greatest utility, that is, those producing the "greatest good for the greatest number."

But Mill and Bentham didn't just write these ideas, they lived them. When John Stuart Mill, James's eldest son, was a toddler, his father and Jeremy Bentham devised a unique social experiment, based on utilitarian principles, with John as its subject. In utilitarianism, all differences in the character and abilities of people are thought to result from differences in their lives. The birth of John Stuart afforded these men an opportunity to test this assumption. If they could give John Stuart the right kind of education, they reasoned, they could shape him into a child prodigy, an independent thinker, a world-class philosopher, and a champion of utilitarianism. John's life thereafter was given over to this idea.

To produce such clear-cut results, John's education would have to start early and be thorough and rigorous. Since James Mill worked at home writing, he became John's tutor, devoting three or four hours a day to the task. John's daily routine began with studies at six in the morning. He would stop for breakfast at nine, then study five hours more. Late in the afternoon, he and his father would walk and discuss the books that John was reading. After dinner, lessons in mathematics began. John had no holidays, and when not studying himself, he taught his eight younger brothers and sisters.

Given John's early training, steeped in utilitarian philosophy and the empirical method, it is understandable that his writings would focus on empirical procedures for evaluating the impact of environmental changes. After all, this idea was central to utilitarianism. Also, John had been fed a daily diet of empiricism ever since he could remember—in the books he was given to read, in the conversations that took place in his home, and in the aim of his educational regime.

And given the restrictions of his childhood, his fascination with ideas of control comes as little surprise. Moment by moment, day by day, year by year, John's life had been regulated according to the precise specifications of his father's experimental design.

Despite the toll of this regime on John (he had no normal childhood and in his adult life suffered from depression), James Mill probably judged the experiment a success. John was a child prodigy. He studied Greek at 3, Latin at 8, and read extensively in these languages by the time he was 12. By 12, he also had studied mathematics through differential calculus. Although John never attended college (sending him to Cambridge was considered but dismissed by his father as a waste of time), he later became the great philosopher and proponent of utilitarianism that James Mill and Bentham had envisioned at the start of their bold experiment.

3.3 Mill's Methods: A Definition Of Control

John Stuart Mill was an empiricist. In true scientific spirit, he looked to experience as the final authority for beliefs. Mill was interested in developing rules of procedure that would allow scientists to establish proof of cause-effect relationships in nature. The four methods he devised were, in his view, methods for discovering and proving laws of causation. They specify the precise observations that must be collected in order to reach a clear-cut conclusion that one event, event A, is the cause of another event, X. In Mill's terminology:

> Laws of causation specify an inevitable sequence between two different events, the antecedent and consequent events—the cause and the effect.

As we will see, Mill's methods are ingenious strategies for testing predictions from the hypothesis that event A causes event X, while *controlling for rival hypotheses,* for example, the hypothesis that event B or C is really the cause of X.

In Chapter 1, we discussed several phases in the research process, from selecting a research problem to drawing conclusions from results. In Mill's analysis, the process of discovering and verifying causation was broken down into two general steps: looking for patterns and verifying relationships. We will discuss each of these in turn.

3.3.1 Two Steps of Research

3.3.1.1 Step 1: The event analysis.

In Mill's view, the process of discovering and verifying causal relationships begins with relatively unstructured observation of the phenomena of interest, a process that today would be called naturalistic observation. At first, no patterns are discernible—no links between antecedents and consequences are evident. Instead, as Mill wrote:

> *The order of nature, as perceived at a first glance, presents at every instant a chaos followed by another chaos. We must decompose each chaos into single facts. We must learn to see in the chaotic antecedent a multitude of distinct antecedents, in the chaotic consequent a multitude of distinct consequents (Mill, 1843/1973, p. 379).*

Mill called the analysis of a situation into its component parts the *event analysis.* In his opinion, general rules for the event analysis could not be specified, since the nature and extent of the analysis would depend on the problem being investigated. The only preparation for this phase, in his view, would be a general preparation of the mind, "for putting it into the state in which it will be most fitted to observe, or most likely to invent" (Mill, 1843/1973, p. 380). Mill believed that the event analysis was complete when the researcher could select from the chaos particular antecedents to relate to particular consequents.

Franklin's commissioners, for example, began their work by observing the public magnetism sessions and developing some ideas of the elements that were involved. They decided to discontinue observing the public sessions when it became apparent that "too many things are seen at once for any one of them to be seen well" (Report, 1784, p. 25). At that point, they isolated individual subjects and selected the antecedents and consequents they would study. After considerable thought, they decided to investigate only the immediate effects of the magnetic treatment, administered with and without suggestion; they would look at the patients' reports of their sensations and any unusual behaviors on their part, such as vomiting, convulsions, or fainting. They were not interested in their subjects' personalities, intellectual abilities, or styles of dress.

This first step in research is important. If any part of the event analysis is incorrect or incomplete, important antecedents or consequences can be overlooked. Analyzing the events in a complex situation and deciding on a fruitful hypothesis to test remain part of the creative challenge of doing research. Since any research situation is extremely complex, Mill was well aware of how difficult the event analysis can be; he believed "to do this well is a rare talent" (Mill, 1843/1973, p. 380). Yet skill in observing and analyzing can be acquired. In part, such skill comes through training in one's discipline; in part, by becoming familiar with the analyses of past researchers working on similar problems. Reading the published research is an invaluable aid to clearly identifying the important elements to consider in a particular area of research.

3.3.1.2 Step 2: The research design.

In Mill's second step, the researcher finds out which antecedents are connected to which consequents, either selecting a particular antecedent and observing what follows it or selecting a particular consequent and determining what precedes it. Mill called this step "varying the circumstances." Varying the circumstances means observing the phenomena you are studying under varying conditions.

The varying observations could be made in either of two general ways: either *by finding* naturally occurring variations or *by making* the variations by artificial arrangements. Recall that these are the two general types of research design discussed in Chapters 1 and 2: the *observational study* is based on naturally occurring events, the *experiment* on artificial arrangements. Mill saw no logical difference between the results of research using these two approaches— "as the uses of money are the same whether it is inherited or acquired" (Mill, 1843/1973, p. 381). (Today, we make a sharp distinction between these two approaches; see Chapters 5 and 6.)

Mill believed that, for the study to lead to conclusive results, the variations in circumstances that we find or make should follow the form of one of four methods: the *method of difference, concomitant variation, agreement,* or *residues.* In Mill's view, all research strategies for determining cause-effect relationships can be reduced to these four. Which method to apply depends on the particular problem being studied and how much is known about it. Since the method of residues is rare in psychological research, we will skip this one and discuss the other three, starting with the method of difference, the one that Mill considered the best.

3.3.2 Method of Difference

3.3.2.1 Case: The magnetized seamstress.

One of Franklin's experiments fits the method of difference so well that we will use it as an illustration. Remember the case of the magnetized seamstress, Experiment 1 in the first chapter. The subject was a patient of Deslon, who was known to have convulsions when exposed to the standard magnetic treatment. When she knew she was being magnetized, she regularly had convulsions. When the commissioners gave her the magnetic treatment without her awareness by magnetizing her through a paper screen, she had no convulsions. From these two observations, that the standard treatment given with the woman's awareness caused convulsions, and that the treatment given without awareness did not cause convulsions, the commissioners were able to conclude that suggestion rather than the magnetic treatment caused the convulsions.

Because the *only difference* in antecedents for these two observations was the suggestion to the subject that she was being treated, the only explanation for the convulsions occurring in the first observation but not in the second is suggestion. It couldn't be the magnetic treatment that caused the convulsions because the patient was treated in the second observation but no convulsions occurred.

3.3.2.2 General research design.

This experiment allowed the researchers to reach a clear-cut conclusion because of its design, which contrasts two observations differing only in a *single* antecedent. The general form of this design, called the *method of difference*, can be clearly stated using a diagram.

Let "X" indicate the occurrence of the consequent event under investigation, and "no X" its absence. Also, let the letters "A," "B," "C," and so on indicate different antecedent events and "no A," "no B," and so on, the absence of these events. With this notation, we can state the method of difference as follows:

Method of Difference
If we can find or make two observations that have the following form? Obs 1: no A + B + C + no D + etc. → no X Obs 2: A +B + C + noD + etc. → X *Then* we can conclude that antecedent event A is the cause of consequent event X.

The diagram shows one observation where X occurs and one where X does not occur. The only difference in antecedents is that event A is absent in Obs 1 and present in Obs 2. Since no other antecedent changes from Obs 1 to Obs 2, we can conclude that event A causes event X.

According to the method of difference, we can prove A to be the cause of X if our observations are:

1. *consistent* with A being the cause of X and
2. *inconsistent* with *any other event* causing X.

If we can rule out *every possible cause other than A,* then, by elimination, A is the cause of X.

The observations in the diagram are consistent with A causing X. In Obs 1, A is absent and X does not occur, and in Obs 2, A is present and X occurs.

The observations also are inconsistent with any other event causing X. Take B, for example. Could B be the cause of X? In Obs 1, B occurs but X does not occur, so B does not cause X.

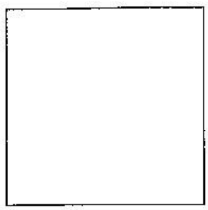


FIGURE 1

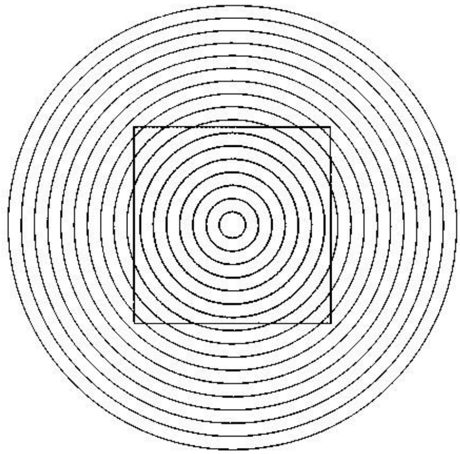


FIGURE 2

To understand the method of difference better, consider the following experiment on visual perception. Look at the square in Figure 1. Now look at Figure 2. The sides of the rectangular figure bow inwards. But the bowing of the sides is an illusion! The figure actually is square; if you are not convinced, place a straight edge next to the lines in Figure 2. The only difference between the drawings in Figures 1 and 2 is the background of circles added to Figure 2. This single difference, then, in accord with the method of difference, is the cause of the visual illusion.

3.3.2.3 Definition of control.

Mill's method of difference is the source of an important principle of research design that we presented in Chapter 1:

> To the extent possible, the researcher tries to eliminate potential rival explanations of the results by holding the conditions associated with them constant. This strategy is called *controlling for rival hypotheses.*

Mill's method provides us with a definition of *control* in a study. The events B, C, D, etc. in the diagram of the method are said to be controlled. If an event is held constant, it cannot cause changes in other events and thereby influence the results of the study. With good controls, the study leads to clear-cut conclusions.

> In a *well-controlled study,* or a study having *good controls,* the experimenter succeeds in controlling for rival hypotheses.

A study which fails to control for critical events is said to be *confounded.*

> In a *confounded study,* the effects of the antecedent of interest are mixed up or confused with uncontrolled events, allowing for rival explanations of its results.

Mesmer's demonstrations of cures through magnetic treatment are good examples of confounding. Mesmer saw his cures as the best evidence in favor of his theory of animal magnetism. But the possibilities of spontaneous cures or cures through suggestion could not be ruled out; they were confounded with the magnetic treatment.

3.3.2.4 Control using a placebo.

Spontaneous cures and the effects of suggestion are still major concerns in experiments evaluating the effectiveness of treatments for illness, and researchers still use Franklin's strategy of giving the patient a placebo treatment. Studies evaluating new drug treatments, for example, compare observations of patients who receive the drug with observations of patients who receive a *placebo.* The placebo looks like, tastes like, and even has the same side effects as the experimental drug, but it is missing the drug's active ingredient. Since all patients receive a dose of something, all receive the suggestion that they are being treated. In this design, spontaneous cures should occur equally in both conditions and thus not favor the active treatment.

The most sophisticated version of this experiment is a *double blind study.*

> In a *double blind study,* neither the patients nor the observers know until the end of the study which patients got the placebo and which got the active drug.

The patients are kept unaware to avoid the effects of suggestion and the observers in order to eliminate any possible influence of their expectations either on the observations they make or on their treatment of the subjects. Logically, then, any difference in the improvement rates of the groups can be attributed only to the active ingredient given to one of the groups, since this is the only antecedent that differs between them.

3.3.3 Method of Concomitant Variation

The method of difference is particularly well suited for designing experiments in the laboratory where the experimenter has control over relevant aspects of the situation. The method of difference assesses the effects of the *presence* or *absence* of an antecedent on the presence or absence of a consequent event: A subject is magnetized or not magnetized; she either does or does not have convulsions.

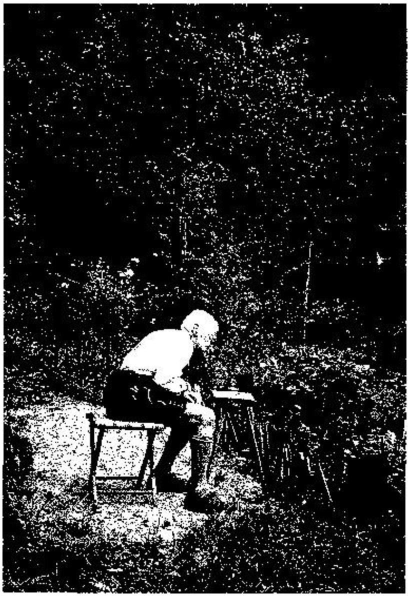
In some investigations, however, the experimenter cannot control the antecedent events, and sometimes the consequent events are not either/or events, as required by the method of difference. Mesmer started his investigations by wondering what effect the moon had on his patient's symptoms.

He certainly could not study the influence of the moon by presenting and removing it. But he could study how variations in the position of the moon would be associated with variations in his patient's symptoms. He wanted to look not only at the presence or absence of the consequent events, his patient's symptoms, but at their intensity and at any other qualitative changes in them. Mill introduced the *method of concomitant* variation to cover this type of situation.

3.3.3.1 Case: The dancing bees.

A study on bees by Karl von Frisch (1950) provides a good illustration of the method of concomitant variation. As in Mesmer's study, the antecedent and consequences of interest to von Frisch were not the either/or events required by the method of difference.

Von Frisch observed that soon after a honey bee discovers food and returns to its hive, hundreds more bees from the hive arrive at the newly found food. Von Frisch wanted to know how the bees so quickly found food, which was sometimes very far from the hive. He hypothesized that the first bee communicated the location of the food to the other bees. To test this hypothesis, von Frisch decided to study what a returning bee did and how the other bees responded.



Karl von Frisch observing bees in his garden.

He designed special observation hives and developed an ingenious system for applying dabs of paint to individual bees so that he could identify each bee in a hive. Von Frisch found that when the bee who discovered food got back to the hive, she transferred the food from her stomach to other bees and then performed a dance. Von Frisch was able to distinguish two variations in the dance: in the "round dance," the bee was "whirling around in a narrow circle, constantly changing direction"; in the "wagging dance," there was "very striking, rapid wagging of the bee's abdomen" during one part of the dance. Von Frisch also observed that the wagging was done at different rates, from about 8 to 36 wags per minute. After the dance was performed, the other bees became excited, and soon after many of the bees would arrive at the distant food source.

Von Frisch first ruled out the possibility that the bees simply followed the leader. Then he came up with the hypothesis that the type of dance the bees did communicated the distance of the food to the other bees. The method of difference, which calls for variation in the presence or absence of the antecedent, was not appropriate to test his idea. Instead, von Frisch chose to vary the *distance of the food* from the hive to see if the *dance* varied systematically with the change in distance. This type of variation of events, in degree or in kind, is the focus of Mill's method of concomitant variation.

3.3.3.2 General research design.

The diagram below shows that this method is similar in its logic to the method of difference; the difference is that the presence and absence of events is replaced here with variations in degree or changes in quality, as indicated in the diagram by the subscripts. Following the logic of the method of difference, the observations in the diagram are consistent with the conclusion that the variation in A causes the variation in X. Other explanations of the change in X are ruled out because only A changes between observations 1 and 2; all other events are controlled.

Method of Concomitant Variation
If we can find or make two observations that have the following form: Obs 1: A_1 + B + C, + not D + etc. > X_1 Obs 2: A_2 + B + C, + not D + etc. > X_2 *Then* we can conclude that the *variation* in event A, from A_1 to A_2 is the cause of the *variation* in event X, from X_1 to X_2.

Applying the logic of the *method of concomitant variation*, von Frisch systematically changed the distance of the food source from the hive and observed the effect on the dance. He found that when the food was less than 50 meters from the hive the returning bee did the round dance. Between 50 and 100 meters from the hive, the round dance changed into a rapid wagging dance. Then from 100 up to 6,000 meters (3.7 miles), the rate of the wagging decreased systematically with the distance of the food from the hive. A slower rate of wagging was associated with greater distance. The graph above shows von Frisch's remarkable results for 3,885 observed dances of bees returning from food sources located from 100 to 6,000 meters from the hive. The graph plots the turns per minute versus the distance of the food from the hive in meters.

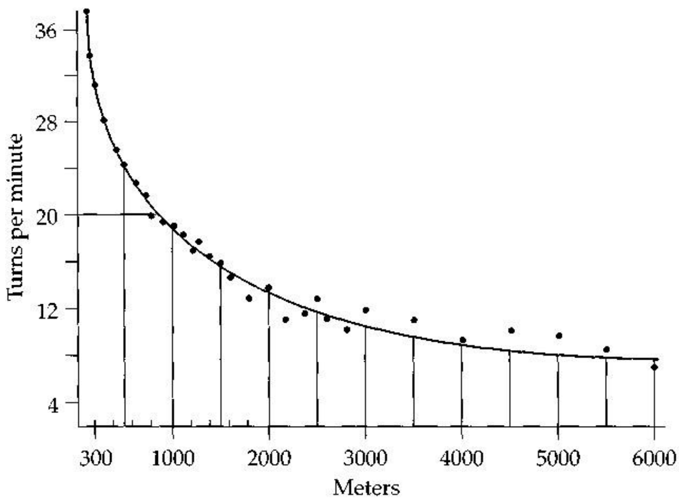


FIGURE 3 The distance of the feeding place is indicated by the number of turns of the wagging dance within a given time (adapted from von Frisch, 1950).

From these results, von Frisch concluded that in the language of the bees the round dance means food very close to the hive and that a decreasing tempo of the wagging dance indicates food progressively farther away, up to a distance of almost 4 miles! In additional experiments, von Frisch discovered how bees communicate the direction of food from the hive; they do their dances at different angles to the sun!

In terms of the diagram of the method, the antecedent event (A) that was varied in this experiment was the distance of the food from the hive. Numerous locations, $A_{(}$ - A_n, were used in the experiment. As far as we know, all conditions other than the distance from the hive of the nectar (e.g., the type of feeding dish, sugar concentration, scent of the nectar, etc.) were kept constant. The consequent event (X) was the type of dance the bees performed; the dances varied depending on changes in A. The observations were consistent with the hypothesis that varying the distance of the food resulted in the different dances. Because the bees who saw the dances were able to find the food, von Frisch concluded that the dance variations communicated the distance of the food. Other rival explanations of the results could be ruled out because conditions other than A were controlled.

The method of concomitant variation is a basic tool in modern research. Many modern studies are concerned with the effects of variations in degree and kind. Studies of perception show that differences in the wavelength of light result in different colors being perceived; drug studies investigate how types of side effects are related to the doses of drugs; developmental psychologists study how different styles of thinking are typical of different stages of development.

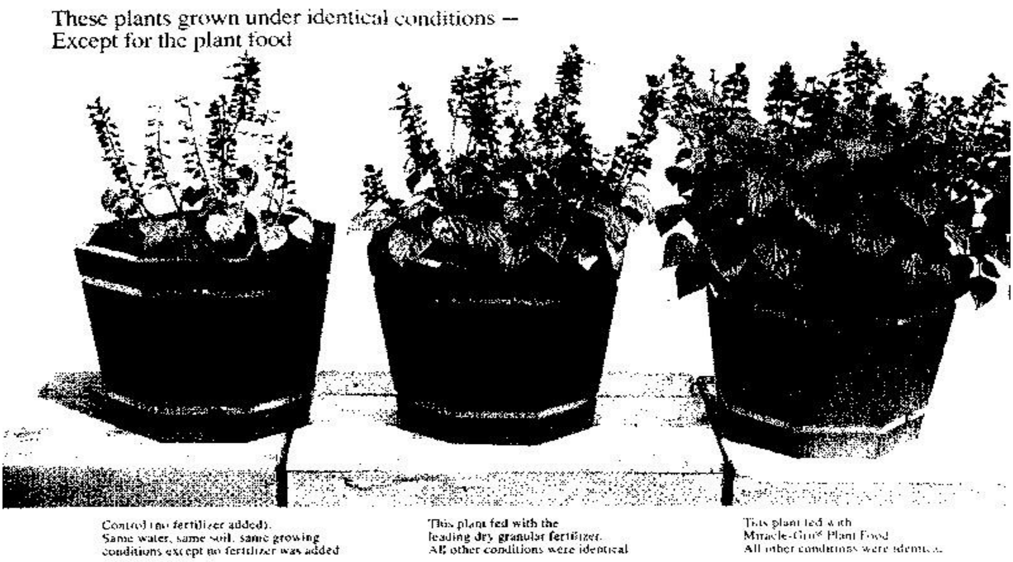


FIGURE 4 Advertisement relying on the Method of Concomitant Variation

The logic of the method of concomitant variation has become so familiar that it is now part of our popular culture. The advertisement in Figure 4 uses the method to demonstrate how plant growth is related to the type of fertilizer used in cultivation.

3.3.4 Method of Agreement

3.3.4.1 Case: A mysterious disease.

The methods of difference and concomitant variation are suited to research where the experimenter tests a well formulated hypothesis about the effect of a particular antecedent. Sometimes, though, the experimenter is faced with an effect and needs to discover its cause rather than to test whether a hypothesized event is the cause. A dramatic illustration occurred in Philadelphia in July 1976, when an unknown disease suddenly struck down a number of people. Health officials needed to discover its cause immediately to prevent further outbreaks and to help in finding a cure. The methods of difference and concomitant variation were of no help because the antecedent (A) of the disease (X) was unknown.

3.3.4.2 General research design.

The health officials began at once to apply Mill's *method of agreement*. All known cases in whom the symptoms (X) occurred were examined to determine whether they shared any antecedents in common. According to the method of agreement, the *single* antecedent common to the cases is the cause of the consequent, X.

Method of Agreement
If we can find or make observations that have the following form: Obs 1: A + B + no C + no D + etc. > X Obs 2: A + no B + C + D + etc. > X *Then* we can conclude that antecedent event A is the cause of consequent event X.

X occurs in both observations and the observations have only one antecedent in common, A. Since the observations are consistent with A being the cause of X and inconsistent with any other antecedent causing X, A must be the cause of X.

When the health officials interviewed the victims, as you might expect, they found many personal and background characteristics that were not common to the victims (e.g., their occupations, home towns, where they ate, their contact with pigs—the officials suspected swine flu), and the general condition of their health; these characteristics were eliminated as potential causes of the disease. The investigators discovered, though, that many of the victims were attending an American Legion convention and staying at the same hotel, the Bellevue Stratford, or had walked past the hotel. The hotel then became the focus of an intense investigation. An examination of the premises uncovered no known agent, no virus, bacteria, or toxin, that could be the cause. Nevertheless, the hotel soon closed due to poor business.

Biological work in the laboratory finally revealed an unknown microorganism in blood specimens taken from the victims. Additional tests on the blood samples yielded results that were consistent with the conclusion that the microorganism caused the disease. The bacterium, later named legionella, apparently grew in the water towers on the roof of the hotel. Water vapor containing it had infected passersby on the street below, and fans had blown the vapor into the hotel (Astor, 1983).

This case illustrates how the *method of agreement* is used in the initial stages of research to find common antecedents to a phenomenon of unknown origin. These antecedents suggest hypotheses that can then be put to more rigorous tests using the *methods of difference* or *concomitant variation*. The method is commonly used in just this way to identify the causes of disease and the toxic agent in cases of food poisoning. For example, it led to the discovery that several persons who died suddenly in Chicago in the fall of 1992 had all been poisoned by cyanide-laced Tylenol capsules. Recently, researchers used the method to suggest a specific genetic basis for homosexuality by showing that several sets of homosexual brothers shared genetic material on the X chromosome (Pool, 1993).

3.4 The Language of Variables

In Mill's methods, the required observations are specified in terms of events being absent or present, or present to a certain degree. Researchers now speak of *variables* rather than events.

A *variable* is a classification or measure of the properties of people, animals, objects, or events, (e.g., height, temperature, sex, intelligence, anxiety, etc.).

3.4.1 Independent and Dependent Variables

Instead of referring to antecedent and consequent events, we now refer to *independent and dependent variables.* The modern terms were borrowed from mathematics. In a mathematical formula, such as $y = 2x + 7$, the variable y is called the *dependent variable,* because its value *depends* upon the value of x. For example, if $x = 1$, then $y = 9$; or if $x = 2$, then $y = 11$. The value of *x,* by contrast, can be set to any value *independent* of the variable y.

In an experiment, the situation is analogous to that in mathematics. The experimenter sets the values of the independent variable, either by creating special conditions or selecting the conditions in nature, and then observes whether or not the variation on the dependent variable is related to the variation on the independent variable.

For example, in his study of bees, von Frisch placed food at several distances from the hive and watched the dance the bees did when returning from the food. He observed that the type of dance done by the returning bees, the dependent variable, depended upon the distance of the food from the hive, the independent variable. In Franklin's study of the magnetized seamstress, the independent variable was the presence or absence of suggestion, and the dependent variable was the presence or absence of convulsions.

In general terms:

An *independent variable* is a variable that underlies the creation or selection of the antecedent conditions in a study.

A *dependent variable* is a variable suspected to be a consequent of the independent variable.

3.4.2 Quantitative and Qualitative Variables

Mill's distinction between events that occur in degrees or in their presence or absence is reflected today in the distinction scientists make between *quantitative* and *qualitative* variables.

> A *variable* is quantitative if its possible values fall on a *numerical continuum;* the values represent the degree or amount of the dimension being measured.
>
> A *variable* is *qualitative* if its possible values differ in *kind or quality* rather than in degree.

Height, weight, degree of extroversion, degree of depression, SAT scores, or the amount of a drug are all quantitative variables. Sex (male vs. female), political party (Democrat, Republican, or Independent), or type of treatment (drug vs. placebo) are qualitative variables.

This change in terms, from *events* to *variables,* is more than just a change in vocabulary. Mill chose the terminology of events because this was a natural way of expressing causal relationships, for example, event A causes event B. However, psychologists often wish to study other types of relationships, relationships of correlation, in addition to those of cause and effect. For example, a psychologist might test whether scores on a test reflect brain damage, or if certain childhood experiences increase the risk of adult disorders. Because the language of variables is general, it is convenient for expressing any type of relationship.

Stated in terms of variables, both Mill's method of difference and his method of concomitant variation require that observations on the dependent variable be made at different values (or levels) of the independent variable, while controlling for all other variables. The conclusion, given observed differences on the dependent variable, is that the variation of the independent variable causes the variation in the dependent variable. In the language of variables, both methods can be stated the same way, so in the discussion to follow, we will refer to both simply as Mill's experimental method.

3.5 LIMITATIONS OF MILL'S EXPERIMENTAL METHOD

Mill made a strong claim for his experimental method. He thought that, if properly applied, it could *prove* laws of causation. That is, Mill believed that the method could establish, for all occasions and for all times, that a law of causation is true. We now realize that experiments cannot offer such proof. The problems lie in three aspects of the method:

1. The problem of controlling for all variables other than the independent variable—the question of the *internal validity* of the study;

2. The extent to which the conclusions of the study generalize to other circumstances—the question of the *external validity* of the study; and

3. The extent to which the independent and dependent variables are valid measures (or operational definitions) of theoretical concepts—the question of the *construct validity* of the study (Cook and Campbell, 1979).

3.5.1 Internal Validity

Today, if a study is well controlled, it is said to be internally valid.

> A study is *internally valid* to the extent that important variables are controlled, so that the actual relationship between the independent variable and the dependent variable can be observed.

Mill's experimental method requires that all variables other than the independent and dependent variable be controlled. If this could be achieved, any change in the dependent variable would have to be caused by the change in the independent variable; there simply would be no other explanation.

However, the complete control called for by Mill's method is an unattainable ideal. In actual studies, it is impossible to control for every single thing: Mesmer overlooked suggestion; von Frisch initially overlooked distance and thought that the bees' dances indicated the type of food they had found. All experimenters can hope to do is to control for variables known to be important. So, there is no way to *guarantee* that a study is internally valid.

It is not uncommon for pioneering studies on a topic to overlook variables that researchers routinely control in later stages of the inquiry. A good example of this is McGinnies's experiment on "perceptual defense" (McGinnies, 1949). Based on psychoanalytic theory, McGinnies hypothesized that threatening visual stimuli evoke defensive responses by the perceptual system to prevent the stimuli from being consciously recognized. He believed that these defenses prevent the anxiety that would accompany the recognition of threatening stimuli.

In his experiment, threat words, such as "whore" and "penis," and neutral words, like "stove," were presented to women college students for very brief exposure times using a tachistoscope. A tachistoscope is an apparatus that allows the researcher to present visual stimuli for a controlled period of time. The exposure times were gradually increased until each word could be identified correctly. McGinnies predicted that, because of perceptual defense, the threat words would require greater exposure times to be correctly identified. His prediction was confirmed.

This study appeared to be well controlled. Each subject was exposed to both threat and neutral words, to control for visual acuity. The lighting for both types of words was identical; the words were presented in the same size print; the threat and neutral words averaged the same number of letters. Nevertheless, the study immediately was criticized for not controlling for a critical variable. Any guesses?

Howes and Solomon (1950) pointed out that McGinnies did not control for word frequency, that is, for how often a word is encountered in reading and conversation. They presented evidence that common words, like McGinnies's neutral words, are more easily identified than infrequent words, like the threat words. McGinnies's interpretation of his results was put in doubt by this alternative explanation. Subsequent studies would have to control for word frequency.

3.5.2 External Validity

Even if Mill's ideal of complete control could be attained, the conclusion of his experimental method would still have to be seriously qualified:

> It is possible that a relationship between the independent and dependent variables exists only for the precise conditions present in the study.

Von Frisch used sugar water in his studies. The type of food was controlled, the same food being used at the different distances. Given just these observations, it could be that von Frisch discovered only the dance that bees do when returning from locating sugar water. Perhaps bees use a different method to indicate the distance of, say, clover, to other bees.

> A study is *externally* valid to the extent that the relationship observed between the independent and dependent variables generalizes to circumstances other than those in the study. This would include generalizing to different types of subjects, different settings, and different types of measures.

Virtually all experiments today follow the pattern of control specified in Mill's experimental method. But the very act of controlling raises issues of generalization. For this reason, experiments have a serious problem with external validity. Do the results of studies done in laboratories generalize to other settings or other subjects? Does saccharin, shown to cause cancer in laboratory animals, also cause cancer in people? Do the results of psychotherapy evaluation studies, in which subjects are carefully selected and the therapy follows a standard protocol, generalize to patients being seen in private practice?

These are difficult questions to answer. It is clear that a single study can do little to establish the external validity of a finding; what is needed is a series of studies that vary in circumstances. Philosophers of science, like Hempel (1966), recommend that hypotheses be replicated under the widest possible range of circumstances for just this reason.

3.5.3 Construct Validity

> The construct validity of an independent or dependent variable is the extent to which the variable is a valid measure of the theoretical construct (i.e., concept) it is supposed to measure (Anastasi, 1988).

If a variable measures something other than the construct intended by the experimenter, then the conclusions of the study may be wrong. For example, what Mesmer took as a valid definition of the construct "magnetic treatment" also could be understood in terms of the construct "suggestion;" these two constructs were confounded in his studies.

Mesmer's difficulty in understanding the true nature of his treatment raises an issue for all researchers—the possibility that an independent or dependent variable thought to measure a single construct actually is complex. In such cases, although researchers may get the results they predict, this may happen for entirely different reasons than they hypothesize.

The perceptual defense study illustrates this problem. Imagine that McGinnies had controlled for word frequency and still found the predicted relationship between the type of word, threat versus neutral (the independent variable), and exposure time (the dependent variable). Would this finding establish the reality of perceptual defense? Not necessarily. The same result could occur due to factors having nothing at all to do with perceptual defense, as Howes and Solomon (1950) suggested.

Howes and Soloman speculated that McGinnies's findings resulted from social factors rather than perceptual ones. It is possible that the subjects recognized the threat words as quickly as the neutral words but were reluctant to report them until they were certain that they were correct. Imagine how embarrassing it would be to announce "penis" and then find out the word that had been presented was actually "genius." To test this explanation, they suggested varying the social circumstances (e.g., having a female experimenter and male subjects).

The issues of internal validity, external validity, and construct validity make it clear that experiments can never *prove* a relationship between independent and dependent variables or guarantee the correct interpretation of one. In fact, the language used by modern researchers for stating their conclusions reflects the fact that the results of a study always are open to question and always require further confirmation. Read a modern study and you will see the conclusions couched in terms like "these results *support* the theory that. . . " or these results *suggest* that. . .," or "the results *confirm the hypothesis* that. . . ." Scientists do not say that their results *prove* a theory true or false.

3.6 APPLYING THE METHODS.

The two basic experimental designs used in psychology today are based on Mill's experimental method. Remember, this method requires that observations on the dependent variable be made at different levels of the independent variable, while controlling for other variables.

> In a *within-subjects design,* each subject is measured on the dependent variable at all the different levels of the independent variable.
>
> In a *between-subjects design,* each subject is measured on the dependent variable at only one level of the independent variable. Different subjects are measured at each level of the independent variable.

Franklin's studies, the studies by von Frisch, and McGinnies's perception experiment were all within-subjects designs. Franklin's group observed the same subject's responses to different treatments. Von Frisch observed the same bees returning from different distances. McGinnies's subjects were presented with both the threat words and the neutral words. By contrast, many studies evaluating psychotherapy and drug therapy use between-subjects designs. Each subject is given only one type of therapy, so there are different subjects in the different treatment groups.

In planning a study, you must choose whether you want to use a within-subjects design or a between-subjects design. In an influential analysis, Campbell and Stanley (1963) compared how these designs handle major threats to internal validity. We will follow their analysis, using a classic research problem, transfer of training, to illustrate the comparison. Since neither design proves to be ideal, the choice between them depends on the specifics of the hypothesis being tested.

3.6.1 Within-Subjects Design

Transfer of training refers to the transfer of skill acquired on one type of task, Skill A, to another task, Skill B. Students are taught Latin because it is assumed that this training will improve their understanding of English. Students study geometry to better their problem solving. Baseball players cross-train lifting weights to improve their hitting. The Head Start preschool experience is designed to provide disadvantaged children with cognitive skills that will be advantageous to them throughout their lives.

The first studies on transfer used a within-subjects design. The subjects were measured on Skill B to determine their level of performance, then trained on Skill A, and finally measured again on Skill B. Several subjects were observed, each receiving exactly the same instructions and procedures.

Campbell and Stanley (1963) identified five major threats to the validity of this type of study: the *threats to internal validity* of *history*, *maturation*, *testing, instrumentation*, and the *threat to external validity* (or construct validity; Cook and Campbell, 1979) of *carryover effects*.

3.6.1.1 History and maturation.

In within-subjects designs, if subjects' scores change from the pretest to the posttest, the change in scores is attributed to the training on Skill A. However, if events other than the training occur between the two measurements, the effects of these events may be confounded with the effects of the training.

The uncontrolled events can occur either in the environment, the threat of *history,* or within the subject, the threat of *maturation.*

> *History* refers to environmental events, other than those associated with the independent variable, taking place between the measurements in a study.
>
> *Maturation* refers to "processes within the [subjects] operating as a function of the passage of time per se (not specific to the particular events), including growing older, growing hungrier, growing more tired, and the like."

In the transfer of training study, if the training phase of the study takes a long time (say, a semester in school), *history* and *maturation* become serious threats. One, or several, of the other activities that subjects engage in over the semester could improve performance on Skill B (the threat of history). Also, at the end of the semester, because they are several months older, the subjects might perform many skills at a higher level than they did at the beginning of the semester, due to maturation.

3.6.1.2 Testing and instrumentation.

In a within-subjects design, subjects are measured on the dependent variable at each level of the independent variable. These multiple measures give rise to two additional threats to internal validity: *testing* and *instrumentation.*

> *Testing* refers to "any effects of taking a test on the scores of a second testing."
>
> *Instrumentation* refers to "changes in the calibration of the measuring instrument or changes in the observers or scores used [that] may produce changes in the obtained measurements."

For example, if observers are being used to record behavior, the researcher must make sure that they do not change their methods as the study progresses. Observers may become blasé as they gain experience in the testing situation; or their ratings may change unintentionally as they become more familiar with the experimental design or the researcher's hypothesis. If measurements are done with instruments, they must be calibrated the same way during all phases of the research; if testing with different forms of a test, the researcher must make sure that they are equivalent.

In the transfer of training study, if the measures of skill were done with paper-and-pencil tests, instrumentation would not be a problem, but testing would be a serious threat. People usually do better the second time they take an ability test, even without any intervening training. This effect is a serious threat to the design since the anticipated effect of training on Skill A also is to increase the scores. Here the effect of testing is confounded with the effect of the independent variable.

3.6.1.3 Carryover effects.

In a within-subjects design, each subject receives all the treatments being studied. If the first treatment a subject receives has a lasting effect, the subject's reaction to the second treatment may be affected. For example, in the perceptual defense study, if subjects are given a threat word first, this may increase their anxiety and influence their reaction to the next neutral word. Campbell and Stanley called this threat to external validity *multiple-treatment interference,* or, more simply, *order* or *carryover effects.*

> *Carryover effects* are "likely to occur whenever multiple treatments are applied to the same respondents, because the effects of prior treatments are not usually erasable."

In some cases, *carryover effects* would be so pronounced that the within-subjects design would not be a practical choice. In a study comparing different treatments for an illness, for example, if the first treatment cures the illness, there would be no point in giving the second treatment. A *between-subjects design* would have to be used in such cases, which occur frequently.

3.6.2 Between-Subjects Design

W. H. Winch (1908) conducted the first between-subjects design on transfer of training. His study is famous because E. G. Boring (1954), a historian of psychology, identified it as the first study in psychology to use a *control group design.*

> A *control group design* is a between-subjects design in which one group of subjects receives the treatment (the *experimental group)* and another group of subjects does not (the *control group).*

To study transfer using this design, the experimental group would be trained on Skill A; the control group would not. Once the experimental group finished its training, both groups would be measured on Skill B. If the experimental group scored higher on Skill B, this would be taken as evidence of transfer of training.

Since each subject is given only one treatment in this design, and is observed only once, the between-subjects design avoids the threats of maturation, history, testing, instrumentation, and carryover effects. However, because *different subjects* are observed in each experimental group, two new threats to internal validity are introduced: *selection* and *mortality.*

3.6.2.1 Selection and mortality.

In Winch's study, the experimental group was a group of school children who received training at memorizing lines of poetry, such as Byron's

> She walks in beauty like the night
>
> of cloudless climes and starry
>
> skies. . . .

The control group did not get this training. The training was expected to facilitate memorization of passages from a history text, like the following:

> At the beginning of the eighteenth century Britain was still chiefly an agricultural country.

The problem here is that any systematic differences between the children in the two groups, in ability, interests, motivation, etc., are confounded with the effects of the independent variable. This threat to internal validity is called *selection:*

> *Selection* refers to systematic differences in the types of subjects assigned to the experimental groups.

Winch tried to equate the groups on their ability to memorize, but who is to say if he was successful? Also, since he personally assigned the children to the groups, there is the possibility that he was biased and unwittingly assigned the best students to the training condition. In addition, one group of students may have matured faster than the other group, the threat of *selection-maturation,* or specific events may have happened to one group and not to the other during the course of the study, the threat of *selection-history.*

Winch's study lasted about three weeks. Because the subjects were children and under the control of their teacher, everyone finished the study. But in longer studies involving less control over the subjects or more demands placed on them, subjects may drop out. If differential dropout rates occur for the groups, and if subjects who drop out are different from those who don't, this would constitute another threat to internal validity, called *mortality* by Campbell and Stanley.

> *Mortality* refers to the differential loss of subjects from the experimental groups.

Mortality is a serious problem. Research evaluating the effects of psychotherapy, for example, can last several months, and when the treatment does not seem to be beneficial, subjects may drop out and seek therapy elsewhere. This is particularly likely to happen in the placebo group, which receives no active therapy.

Selection and mortality present difficult problems for the between-subjects design. If subjects in the experimental groups are different, the effects of these differences can be confounded with the effects of the independent variable. The within-subjects design avoids these problems by observing the same subjects in different conditions of the experiment. When subjects drop out of a within-subjects design, they can be dropped from all the conditions, so there will not be a differential dropout rate.

3.7 Approaches To Dealing with Uncontrolled Variables

Neither the within-subjects nor the between-subjects design offers the perfect control of variables that Mill's method requires. Mill offered no solutions to this problem; but since his day three distinct practical approaches have been developed for dealing with uncontrolled variables.

3.7.1 The Within-Subjects Design with Virtually Total Control

In his now classic research on conditioned reflexes, Ivan Pavlov (1927/1960) the Russian Nobel Prize-winning physiologist, introduced the strategy of handling the problem of uncontrolled variables by regulating virtually all aspects of the experimental situation. His research on the salivary reflex in dogs was carried out in a custom-built laboratory called the "tower of silence." The building was soundproofed to eliminate the influence of uncontrolled stimuli from outside. During the experiment, the animals were observed using a periscope, to avoid uncontrolled interactions between the dogs and the experimenters. Dogs were ideal subjects for this procedure because suggestion posed no threat and because the ease of controlling the animals' lives during and between experiments minimized any threats of history. The experiments on conditioned reflexes also were of short duration minimizing maturation as a threat. The subjects had special surgery so that their salivary glands would discharge directly into a tube outside their mouths. The amount of saliva produced in response to a stimulus was accurately measured, eliminating the threats of instrumentation and testing The experimental design was within-subjects, with only one animal observed at a time (a so-called $n = 1$ design). The results were replicated with other dogs.

The total control achieved by Pavlov is possible only in animal studies, and his approach is well suited for such research. B. F. Skinner (1938) adapted and extended Pavlov's approach to develop the experimental methods used in operant conditioning. Today Pavlov's and Skinner's single-subject designs are used frequently in experiments on animals and people. These designs, as well as the modern extensions of them, are discussed in Chapter 9, Single-Case Experimental Designs.

3.7.2 Statistical Control: Measure the Uncontrolled Variables and Remove Their Effects

In between-subjects designs, it is impossible to control perfectly for differences between subjects in the various groups; subjects may differ in age, education, attitudes, past experience, etc. (the threat of selection we discussed above). However, in some cases, the subjects can be measured at the beginning of the study on variables suspected to be threats to internal validity. Questionnaires can be developed to ask people about relevant past experiences, and standard measures can be used to assess their interests, abilities, and personality traits. Then when the study is over, statistical procedures can be used to remove the effects of these variables on the dependent variable. These variables then are said to be "statistically controlled."

Statistical control is a powerful technique in *observational studies* in which direct control, by holding variables constant, is impossible. Because understanding statistical control requires some knowledge of measurement and the mathematics of correlation, this technique will be discussed in Chapter 5, Correlation, following the presentation of these topics.

3.7.3 Randomized Experiments

In some between-subjects studies, researchers can control the assignment of subjects to conditions. In drug studies, for example, experimenters can decide which subjects to assign to the placebo and which to the active drug. When this is the case, the best approach for control purposes is to *randomly assign* the subjects to the groups.

> In *random assignment,* each subject is assigned to an experimental condition purely by chance, (e.g., by flipping a coin, or drawing the subject's name out of a hat, or by using computer-generated random numbers). This assignment

gives subjects an equal chance of being assigned to any of the conditions of the study.

Random assignment avoids any systematic bias in assignment and permits a statistical estimate to be made of the magnitude of the effects of uncontrolled variables in the study. This estimate of the error due to uncontrolled variables is used in evaluating the results of the study.

The randomized experiment is considered the best available design for experiments because it offers the most powerful method developed so far for overcoming the threats of uncontrolled variables. Unfortunately, it is not always possible to randomly assign subjects to experimental conditions. The logic of randomization, the method of choice when it can be applied, is explained in Chapter 6, Randomized Experimental Designs.

The development of the two approaches to control that use statistics, statistical control, and randomization, occurred in a three-step sequence. First, procedures were developed to measure individual differences. Then, a novel method of analysis was invented to determine the degree of "correlation" between different measures. Finally, these new methods of measurement and correlation were applied to problems in experimental design.

In the next two chapters, we discuss how researchers measure individual differences and determine the correlation between variables.

3.8 Key Terms

Empiricism

Mill’s methods

Utilitarianism

Laws of causation

The event analysis

Varying the circumstances

Methods of difference, concomitant variation, and agreement

Well-controlled vs. confounded study

Placebo

Double-blind study

Variable

Independent vs. dependent variables

Quantitative vs. qualitative variables

Internal validity

External validity

Construct validity

Tachistoscope

Within-subjects design

Between-subjects design

Threats of history and maturation

Threats of testing and instrumentation

Threat of carryover effects

Control group design

Threats of selection and mortality

Virtually total control

Statistical control

Random assignment

3.9 Key People

John Stuart Mill

James Mill and Jeremy Bentham

Karl von Frisch

Thomas D. Cook, Donald T. Campbell, and Julian C. Stanley

McGinnies

Ivan Pavlov

B. F. Skinner

W. H. Winch

3.10 Review Questions

1. Describe John Mill's early childhood and explain how his upbringing was an experiment testing the tenets of utilitarianism.

2. Describe Mill's two steps for establishing laws of causation.

3. Diagram and explain the logic of Mill's method of difference.

4. Analyze the cause of the visual illusion presented in the text using the method of difference.

5. Compare the method of difference and the method of concomitant variation.

6. Explain how von Frisch's study of bees illustrates the method of concomitant variation.

7. Compare the method of difference and the method of agreement. Why are the conclusions of studies that use the logic of the method of agreement often verified by experiments using the method of difference?

8. Explain how researchers used the method of agreement to find the cause of Legionnaires' disease.

9. Why do modern scientists use the language of variables rather than Mill's language of events?

10. Distinguish between independent and dependent variables.

11. What are the three major problems with Mill's experimental method that preclude researchers from reaching certain conclusions using it?

12. State two rival hypotheses to explain the results of McGinnies's experiment on perceptual defense.

13. Identify the major threats to the validity of a within-subjects design.

14. Identify the major threats to the validity of a between-subjects design.

15. Describe three approaches to deal with the threat to validity of uncontrolled variables.