

order to satisfy this criterion, the results in our anthropologist's study would have to favor his hypothesis to at least this degree:²⁴

	Zoco attacks	No zoco attacks
Initiated boys	2	8
Initiated girls	8	2

Conclusion

We have now discussed a number of factors that should be considered in designing hypothesis-testing research in anthropology. On the basis of this material, it is possible to draw up a list of questions that a student of anthropology should ask of any research design. This is true whether the "student" (of whatever status) intends to evaluate his or a colleague's proposed research, or to criticize the published results of hypothesis-testing research. The questions are:

1. Is the research paradigm used the strongest possible one under the circumstances?
2. Do the measurements appear to be satisfactory with regard to their reliability, validity, and precision?
3. Have provisions been made for dealing with the role of chance?
4. Is the execution of the research described in such a way as to be replicable?

If any of these questions cannot be answered affirmatively, there is cause for some concern about the design and, by extension, about the research results that have (or would have) emerged from its use.

²⁴When assessing the results of statistical checks against the rival hypothesis of chance, researchers are "always confronted with the risk of making one of two types of error" (Sellitz et al. 1959:417, italics added). One type of error, known technically as an "alpha" or "Type I" error, is made when a researcher concludes that there are nonchance differences when in fact there are none. A second type of error, known technically as a "beta" or "Type II" error, is made when a researcher concludes that there are no nonchance differences when in fact there are. It can be seen that there is no way to minimize these two types of errors at the same time because they are "the opposite sides of the same coin." That is, the more we raise the level of significance (for example, by raising the level from .05 to .01) to protect against making Type I errors, the more we open ourselves to Type II errors. Correspondingly, the more we lower our level of significance (for example, by lowering the level from .05 to .10) to protect against making Type II errors, the more we open ourselves to Type I errors. The .05 level of significance we have used here has traditionally been viewed as a compromise between these two types of errors, although there are those who believe this is too strict (that is, by using the .05 level we are making too many Type II errors). In any case, even if the anthropologist's results in this example were such that chance could be ruled out, he would still, of course, have to take into account other rival hypotheses to which his design is subject—in this case, selection and selection interaction effects.

3 / Research design: Some anthropological examples

Now that the basic elements of research design have been introduced, we will illustrate their application by turning to examples of research drawn from the literature in anthropology and related fields. This chapter has several interrelated objectives. The primary objective is to elucidate further the strengths and weaknesses of the basic paradigms that were introduced in Chapter 1. In addition, the research reviewed here will illustrate some important techniques that can be used to compensate at least partially for the weaknesses of paradigms such as the static-group-comparison and the pretest-posttest paradigms. This review will also serve to demonstrate that practical measures can be devised for even highly complex and elusive variables of the sort anthropologists must frequently deal with, such as sexual attraction, desire to achieve, tendency to imitate others, and emphasis on compliance in child-training. Finally, this chapter will introduce an important variant of the static-group-comparison paradigm, the "correlational approach."

THE STATIC GROUP COMPARISON IN ANTHROPOLOGICAL RESEARCH

Coming of Age in Samoa

By far the most common research paradigm in anthropology is the static group comparison. While this paradigm is far from the strongest available, the realities of anthropological research are often such that no stronger designs can be employed. A great deal can be learned from the static group comparison if steps are taken to apply it with maximum effectiveness. As will be seen, this paradigm has been applied by anthropologists in a number of different ways, some of which are clearly superior to others.

Let us begin by considering the least satisfactory variant of the static group comparison. This is a version of the static-group paradigm where one of the two groups is conjectural—that is, either the comparison group or the treatment group has no concrete existence but is instead constructed on the basis of the investigator's general knowledge or intuition. This type of static group comparison is frequently

encountered in earlier anthropological works. To illustrate it, let us focus on one classic study—Margaret Mead's *Coming of Age in Samoa* (1928).¹

The research problem of central importance for Mead in this study concerned the effects of adolescence on children. Many social commentators in Western societies had observed that adolescence was a stormy and crisis-filled period. They hypothesized that this was an inevitable result of the physiological changes occurring at this period in a young person's life. Mead doubted the validity of this hypothesis and countered it with an alternative hypothesis of her own. She suggested that the traumatic nature of adolescence in Western society was not the result of physiological changes in the bodies of adolescents, but was due to the culture in which they lived. In her own words (1928:16), "... the question which sent me to Samoa [was]: Are the disturbances which vex our adolescents due to the nature of adolescence itself or to the civilization?" The hypothesis she wished to test, then, may be stated as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Exposure to Western civilization	increases	adolescent trauma.

Mead went to the South Seas island of Samoa to obtain her comparison group. Here she studied a group of adolescent girls who, as they grew up in Samoan culture, enjoyed virtual freedom from the influences of Western culture. Acting as a participant-observer, Mead made highly detailed observations of the emotional and other behavior of a sample of girls. Subsequently, on the basis of these observations, she rated the girls as a group in terms of how much emotional distress they experienced. In her judgment (1928:95), with the exception of a few cases, "adolescence [for these girls] represented no period of crisis or stress, but was instead an orderly developing of a set of slowly maturing interests and activities. The girls' minds were perplexed by no conflicts, troubled by no philosophical queries, beset by no remote ambitions."

Having found a comparison group and having obtained, for the group as a whole, a measure of the dependent variable (adolescent trauma), Mead was faced with the problem of obtaining a treatment group. To do this she resorted to a *conjectural* treatment group. She made no equivalent study of adolescents in West-

¹ We select this example because it is beyond question one of the best-known anthropological classics and because, for the period, it represents research of very high quality. We emphasize this last point because it is important for the reader to realize that the deficiencies of this study from the standpoint of research design are not specific to this one work but are representative of anthropological studies undertaken during this period, when, it might be said, the criteria for adequacy of research design had not yet become fully incorporated into the culture of anthropology. It should be pointed out, however, that the anthropology of this time represented a very significant improvement over the pioneering anthropological work of the late nineteenth century, in that it had come to be standard practice for data to be collected by a process of intensive, first-hand participant-observation, rather than being gleaned from the reports of nonanthropological observers, as was the case in much nineteenth-century work.

ern civilization, but instead relied on her general knowledge of a Western society, American society, to construct a treatment group conjecturally. She rated this conjectural group of American adolescents as suffering from a high degree of emotional disturbance, or in her words, "storm and stress." Having supplied a treatment group—albeit a conjectural one—Mead concluded that her hypothesis had been supported. The adolescents in the comparison group, who were isolated from the influence of Western civilization, showed little or no emotional stress at adolescence. The conjectural American adolescent treatment group, which was exposed to Western civilization, did suffer "storm and stress." The results were as predicted by the hypothesis.

We can diagram Mead's study as follows:

Treatment group (conjectural) (American adolescent girls)	(X O)
Comparison group (Samoan adolescent girls)	O

In this diagram, the parentheses bracketing the two symbols for the treatment group indicate that the group is conjectural. It is also possible, as has frequently been the case in anthropology, for the comparison group to be conjectural. Such a case would be diagrammed as follows:

Treatment group	X O
Comparison group (conjectural)	(O)

Here the parentheses bracket the row corresponding to the comparison group to signify that no real cases in this group were actually studied.

Today, we can still admire Mead's study for its many valuable insights into adolescence and for the quality of its description of Samoan culture. The study may also serve as an example of the weaknesses of the version of the static-group-comparison paradigm that employs a conjectural treatment or comparison group. In addition to the weaknesses common to the standard static group comparison, this version of the paradigm contributes a very serious new error source—namely, the imprecision of a conjectural comparison or treatment group. If the investigator does not actually include *both* the treatment and comparison groups in his research, but, as in Mead's study of Samoan girls, relies on general knowledge to conjecture one of the two groups, the possibility of error is greatly magnified. The most serious question raised by such a procedure concerns the equivalence of the measurement procedures for the two groups.

In the present example, Mead based her measurement of emotional stress on extensive, more or less systematic observations of the behavior of a specific group of Samoan adolescent girls. To rate her conjectural American adolescent group, however, she must have relied on a very different body of information, no doubt including her own memories and experiences as an adolescent member of American culture, her subsequent more or less casual observations of American adolescents,

and various scholarly studies of American adolescents. She did not make a detailed, systematic behavioral study of a specific group of American adolescents comparable to the one she carried out in Samoa (at least, we are not told if she did). Thus, we are forced to the conclusion that since her ratings of emotional stress for American and Samoan adolescents were based on different sorts of evidence and different measurement procedures, the comparability of the ratings for the two groups of adolescents is compromised.

There is another limitation of Mead's measurement procedure that we should examine—she placed very heavy reliance upon herself as an instrument to measure the variables of interest. Again, we would emphasize that this limitation is not peculiar to Mead's study; rather, it is characteristic of much anthropological research. Mead herself acknowledges this limitation in a statement remarkable (for the time) in its sensitivity to the problem. Regarding her judgment of adolescent emotional stress and the other findings of her study, she notes (1928:153): "The conclusions are . . . all subject to the limitation of the personal equation. They are the judgments of one individual upon a mass of data, many of the most significant aspects of which can, by their very nature, be known only to herself."

In a test of a hypothesis, exclusive reliance on the personal judgment of the researcher as the means of measuring key variables is undesirable. Experience has shown that a person's judgment of a situation can be markedly affected by his expectations—that is, a person tends to see in a situation what he expects to see, even if it is not objectively present (see, for example, Rosenthal 1966). Complete reliance on this sort of judgmental measurement by the anthropologist can introduce considerable measurement error. Not only is there some danger of the anthropologist unconsciously "misjudging" various behaviors, but he may also, again quite unconsciously, tend to be somewhat selective in his observations, being particularly alert to behavior that conforms to his expectations and tending to overlook other behavior he does not expect. This is *not* to say that all anthropological observations are inaccurate, but does point up the serious possibility of substantial measurement error which may result from unconscious distortion and/or selection of the data on the part of the anthropologist who relies so heavily on judgments of this type.

The problem of unconscious observer bias can be ameliorated somewhat if the data collected by the anthropologist are rated by independent judges who are not aware of the hypothesis being tested. Also, if several independent judges are used, an estimate of the interrater reliability of the rating system can be obtained. This approach still does not completely eliminate the problem of inadvertent selection of behavior that favors the hypothesis, however, since we are still dependent on the anthropologist himself to observe and record a behavior sample to serve as a basis for ratings.

This problem has not been ignored by anthropologists. In recent years, as they have come to recognize the serious consequences stemming from misjudgment of data and unconscious data selection, there has been a move toward supplementing participant-observation by using more objective data-collection instruments. These devices and procedures include standardized interview protocols and schedules, questionnaires, projective tests, and regularized procedures for sampling and re-

ording behavior which reduce the possibility that an anthropologist's expectations may unconsciously affect his reporting of the data.

But Mead's study is not only weakened by measurement problems. Even if we assume that Mead's conjectures regarding the treatment group were accurate and her measurements of the Samoan comparison group were valid, there are other sources of possible error. For example, as was pointed out in Chapter 1, the static-group-comparison paradigm guards against the effects of extraneous variables only to the extent that the treatment and comparison groups are affected by these variables in the same manner. If there is an interaction between selection and the effects of extraneous variables, that is, if the cases in the treatment and comparison groups are differently affected by extraneous variables, there is a very real danger that an observed difference between the two groups in regard to the dependent variable may be due to one or several of these extraneous factors and may have no connection with the independent variable. It is possible to identify several factors (extraneous variables) that might differentially affect Mead's treatment and control groups and might therefore be responsible for the difference in amount of observed emotional stress.

For example, the genetic makeup of the two groups of adolescents being compared is obviously different. There are strong indications that there are important genetic components in psychological factors such as emotionality (Eysenck and Prell 1951). It is therefore possible the Samoan adolescents were constitutionally less subject to emotional stress. Another possibility is that the tropical illnesses to which Samoan teenagers were subject had the effect of making them somewhat listless and therefore less subject to emotional arousal. These two alternative hypotheses cannot be ruled out, since the factors involved in them cannot be controlled by Mead's application of the conjectural static group comparison; in other words, it cannot be assumed that these factors affect both groups equally.

A static group comparison can be greatly strengthened by selecting treatment and comparison groups that resemble each other as much as possible and that are likely therefore to be affected by extraneous factors in nearly the same manner.² When dealing with a hypothesis such as Mead's, there is little possibility of obtaining comparison and treatment groups that are similar in regard to the major extraneous variables. It would almost certainly be futile, for example, to try and locate

²The goal of selecting societies or cultures for study that are similar in virtually every way except for the one issue or variable of interest is the core of what is known in anthropology as the "method of controlled comparison." Eggan's discussion of this method is one of the most thorough and widely known (Eggan 1954); Nadel's study of witchcraft in four African societies is perhaps the most famous anthropological application of this method (Nadel 1952). Although seldom explained in these terms, the effect of such a method of sample selection is to reduce the likelihood that extraneous variables will affect the two groups differently and in this way protect against a serious rival hypothesis. Such a procedure is desirable when using this method because a "controlled comparison" is essentially the same as a static-group-comparison paradigm. And, as we have noted, the problem of extraneous variables with this paradigm can be minimized if the groups compared can be assumed to have responded similarly to them; this assumption is more tenable if the groups are similar. Unfortunately, however, most instances where this method has been employed by anthropologists have been for the comparison of relatively few cultures or societies (often as few as two) and this has meant that it was impossible to consider the role of chance.

a non-Western cultural group that has the same genetic makeup and the same diet as a Western cultural group. In the case of many other, more specific, hypotheses, however, it is possible to locate similar treatment and comparison groups. For an example of this, let us return to Mead's study.

Mead realized that even if she were able to demonstrate conclusively that Western cultures produced a high degree of adolescent emotional stress, such a conclusion would be of limited practical value since it would provide no indication of what specific cultural institutions might be changed in order to improve the lot of adolescents. She therefore went on to develop some relatively specific hypotheses relating particular cultural features to adolescent distress. She suggested, for example, that when multiple standards of sexual morality are found in a culture (as in American culture) an adolescent is placed under increased stress, since he or she faces the dilemma of which standard to adopt. Samoan culture, Mead suggests, has in effect only one standard of sexual morality. This makes it possible to develop a more specific and potentially more useful hypothesis that can be diagrammed as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Being subjected to competing standards of sexual morality	increases	emotional distress at adolescence.

As we have seen, an attempt to test this hypothesis using a static group comparison between groups that are as different as American and Samoan adolescents runs up against perhaps insurmountable difficulties in that there are many *uncontrolled extraneous variables* (variables which will affect the two groups differently) whose effects can masquerade as those of the independent variable. A more powerful test of this hypothesis might be carried out by selecting for a static group comparison two groups of girls drawn from the *same* culture but differing in the extent to which they are exposed to competing standards of morality. Since both groups would be drawn from the same culture, the problem of extraneous variables would be considerably reduced as the groups would be affected in the same way by many, if not most, extraneous factors. Mead herself pointed out that such a comparison would have been possible in her research setting, since some of the girls went to live in the households of native pastors where they were exposed to moral standards very different from those which they had learned in their own families (1928:155).

In summary, employing the criteria for evaluating a research design that we proposed at the conclusion of Chapter 2, Mead's study can be criticized on at least five grounds. First, the paradigm she used was not the strongest possible under the circumstances. She could at least have utilized the standard static group comparison instead of the inferior conjectural version of the static group comparison. Second, the reliability of the measurements is seriously in doubt due to the facts that (a) the researcher used herself as a measuring instrument for adolescent stress, apparently by a primarily intuitive process, and (b) no direct measurements were

made on the treatment group—a group which was essentially conjectural. Third, the validity of the measurements is seriously in doubt for the same reasons given regarding the reliability of the measures. Fourth, the replicability of the design is seriously deficient, again largely because of the subjective nature of the measurement process. And fifth, the absence of an actual treatment group made it impossible to apply statistical tests to take into account the role of chance.

Child Marriage in Taiwan

In the discussion of Margaret Mead's *Coming of Age in Samoa*, it was suggested that it is possible to strengthen a static group comparison considerably by obtaining treatment and comparison groups which are as similar as possible. This strengthens the paradigm because, in general, the more similar two groups are, the greater the likelihood that they will be exposed to the same extraneous variables. A study which illustrates this principle is Wolf's examination (Wolf 1970) of Westermarck's hypothesis (Westermarck 1926) that childhood association produces "sexual aversion"—in other words, that a boy and girl who are in close association with one another from infancy will develop no sexual interest in one another and indeed may actively feel repulsed by thought of sexual contact. Wolf did fieldwork in an area of Taiwan where the existence of two contrasting types of marriage made it possible to construct a static-group-comparison test of the Westermarck hypothesis. In one type of marriage, which Wolf calls the "minor" form of marriage, the bride-to-be comes into her future husband's household as an infant or small child. She is "adopted" by her future husband's parents and reared together with him until they both reach maturity, at which time their marriage is finalized. In the second type of marriage, called the "major" form of marriage by Wolf, the bride and groom typically do not meet until the time of their marriage, when they are both adults.

As we have noted, Westermarck's hypothesis predicts that intimate childhood association between a boy and girl will result in their having no sexual attraction for one another. The relationship of the variables in this hypothesis may be stated as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Close childhood association	decreases	sexual attraction.

Individuals in "minor" marriages can be considered as cases in a treatment group in regard to this hypothesis since they have grown up in intimate association with their spouse. Individuals in the "major" marriage form provide a comparison group since they have experienced no childhood association with their spouses.

Wolf was able to devise several measures of the dependent variable of sexual attraction. He assumed the number of children produced by a marriage was directly related to sexual attraction between husband and wife. That is, couples who attract one another sexually produce more children than those who are not attracted to one

another sexually. He also assumed that if a marriage ended in divorce or featured adultery involving the wife, this was usually an indication of a low degree of sexual attraction.

Wolf's study may be diagrammed as follows:

Treatment group (couples in minor marriages)	X (childhood association)	O (number of children and outcome of marriage)
Comparison group (couples in major marriages)		O (number of children and outcome of marriage)

The results of this comparison, shown in Tables 3.1 and 3.2, were consistent with the Westermarck hypothesis. Minor marriages showed a much higher incidence of divorce and/or adultery by the wife. Wives in minor marriages also had substantially fewer children.

TABLE 3.1 NUMBER AND PERCENT OF MARRIAGES ENDING IN DIVORCE AND/OR INVOLVING ADULTERY BY WIFE^a

	Minor marriage	Major marriage
Total number of marriages	132	171
Number involving divorce and/or adultery	61	18
Percent involving divorce and/or adultery	46.2	10.5

^a From Wolf 1970:512, Table 5. Reproduced by permission of the American Anthropological Assn. from the *American Anthropologist*, 72(3), 1970.

TABLE 3.2 AVERAGE NUMBER OF CHILDREN AS TAKEN FROM HOUSEHOLD REGISTRATION RECORDS^a

Years of marriage (in five-year intervals)	Minor marriage	Major marriage
1st	1.27	1.81
2nd	1.19	1.62
3rd	1.12	1.54
4th	1.06	1.23
5th	0.54	0.75

^a From Wolf 1970:512, Table 6. Reproduced by permission of the American Anthropological Assn. from the *American Anthropologist*, 72(3), 1970.

Because Wolf obtained his comparison and treatment groups from the same culture instead of from two different cultures (indeed, the groups were both obtained from a single small region in Taiwan) he could reasonably assume that many extraneous factors were the same for both groups. This was not, however, true for all extraneous factors. Importantly, the two groups were exposed to different levels of the extraneous variables of socioeconomic status and adoption. These differences were important because they led to rival hypotheses that might have accounted for the different scores of the groups on the dependent variable.

It could be suggested, for example, that because women in minor marriages were all adopted in infancy this might in itself have been responsible for the smaller number of children resulting from minor marriages. There is some reason to believe that women adopted in this fashion were often mistreated by their foster parents and perhaps did not eat as well as the other members of their adoptive families. Deprivation experienced by such women in their childhoods could have lowered their fertility and caused them to bear fewer children when they married.

Minor marriages also are generally entered into by persons of low socioeconomic standing. This factor might in itself have led to the high rates of divorce and adultery observed for minor marriages. Women, for example, might have been motivated by extreme poverty to leave their husbands in search of a more materially rewarding life situation. Men might have been more likely to desert their wives in order to escape the burden of providing for a family. It might also be suggested that persons of low status may have different standards regarding divorce and extramarital sex. If these standards were more permissive, this in itself could account for the higher incidence of both of these phenomena. The poverty that accompanied low status might also have been responsible for the lower fertility of women in minor marriages, since poor health tends to accompany poverty.

If unrefuted, these rival hypotheses, which follow from the differences between the treatment and comparison groups on the two extraneous factors of adoption and social status, would throw into serious doubt the conclusion that the results of the study support the Westermarck hypothesis. However, Wolf was able to make a strong case that social status and adoption differences were not responsible for the differences in scores on the dependent variable. He did this by matching couples from the major and minor marriage groups on the two extraneous variables. In other words, he selected cases that were equal in terms of these extraneous variables. There were, he found, forty-two women in the major marriage group who had been reared as adopted daughters in preparation for minor marriages. Because of the death of their intended spouses or other circumstances, these women had married in the major form instead. They had, however, been exposed to the effects of adoption, just as were the women who went ahead with minor marriages. They were also, by and large, from the same socioeconomic stratum as the minor marriage wives.

Wolf carried out a new static group comparison. This time, he compared two groups of women—those forty-two who had been "adopted" for minor marriage but had ended up in major marriages, and those who had been "adopted" and did enter minor marriage as planned. In this comparison the effects of the two extraneous variables of status and adoption were essentially controlled since both the treatment group and the new comparison group were affected by them to approximately the same extent. Thus, these two extraneous variables could not account for any dif-

ferences between the two groups regarding their scores on the dependent variable. This new comparison is diagrammed below:

Treatment group (couples in minor marriages)	X (childhood association)	O (number of children and outcome of marriage)
Comparison group (couples in major marriages where the wife was reared as an adopted daughter- in-law in prepara- tion for a minor marriage)		O (number of children and outcome of marriage)

The results of this new comparison were very similar to those of the original comparison. Again, the treatment group of minor marriages, where husband and wife were reared together, produced fewer children and featured much higher rates of divorce and/or adultery involving the wife (see Tables 3.3 and 3.4).

TABLE 3.3 NUMBER AND PERCENT OF MARRIAGES BY ADOPTED DAUGHTERS ENDING IN DIVORCE AND/OR INVOLVING ADULTERY BY WIFE*

	Minor marriage	Major marriage
Total number of marriages	132	42
Number ending in divorce	25	1
Number ending in adultery	42	4
Percent involving divorce and/or adultery	46.2	9.5

* From Wolf 1970:514, Table 8. Reproduced by permission of the American Anthropological Assn. from the *American Anthropologist*, 72(3), 1970.

TABLE 3.4 AVERAGE NUMBER OF CHILDREN BY ADOPTED DAUGHTERS AS TAKEN FROM HOUSEHOLD REGISTRATION RECORDS*

Years of marriage (in five-year intervals)	Minor marriage	Major marriage
1st	1.27	1.78
2d	1.19	1.77
3d	1.12	1.76
4th	1.06	1.31
5th	0.54	0.90

* From Wolf 1970:514, Table 9. Reproduced by permission of the American Anthropological Assn. from the *American Anthropologist*, 72(3), 1970.

Since status and adoption have been effectively ruled out as rival hypotheses accounting for the observed difference in the two groups' scores on the measures of the dependent variables, Wolf was in a much stronger position when he suggested that these results support Westermarck's hypothesis that prolonged and intimate childhood association decreases sexual attractiveness.³

It is likely that the two groups in Wolf's study were differentially affected by at least some extraneous variables other than the two he dealt with by matching. It is not easy, however, to suggest any that might plausibly be held to affect the dependent variable. Nonetheless, Wolf's conclusions may be thrown into contention if he himself or some other researcher is subsequently able to suggest such an extraneous variable.

It is important to note that selection seems to pose little problem in Wolf's research, even though he is using a paradigm which is vulnerable to error from this source. A primary danger in many applications of the static group comparison, as we have seen, is that as a result of selection, the treatment group will already be high (or low) on the dependent variable by the time it is exposed to the independent variable. In such a situation the effect of selection can masquerade as that of the independent variable and lead the researcher into a false inference. This possibility seems remote here, since the research situation is one in which the independent variable, childhood association, typically begins to act not long after birth. It is hard to imagine in what way selection effects could cause the two groups compared to differ on the dependent variable of sexual interest *prior* to the onset of the independent variable. Since much anthropological research deals with independent variables that have an early onset, such as the one here, it is often the case that preexisting differences on the dependent variable due to selection can be discounted as a possible rival hypothesis. (Effects of selection may often continue to be a problem, however. See, for example, the selection-related problem discussed for LeVine's study, pages 42-48.)

Evaluating Wolf's research design according to the general criteria we developed in Chapter 2, we can say the following: The paradigm used is the strongest possible one under the circumstances and has been strengthened substantially by Wolf's efforts to control extraneous variables through the procedure of matching cases. Regarding reliability, two of the measures he employed—the number of children a woman gave birth to and the occurrence of divorce—were taken from population records, which, presumably, were carefully maintained. These data are probably reliable as there is relatively little chance of the researcher misjudging data of this sort. Of course, further evidence of the accuracy of the original documentary records would be desirable. Wolf's information on adultery on the part of wives is perhaps open to question from the standpoint of reliability, in part because his sources for this information were "well-informed" local informants. No interrater reliability checks were reported. (The sensitive nature of the topic probably made it difficult

³ Campbell and Stanley (1966) point out that in some situations matching can introduce serious error due to a statistical artifact known as "regression," although this does not appear to pose a problem in Wolf's research. Such error may be avoided by use of a type of statistical analysis known as analysis of covariance, which achieves the benefits of matching without the risk of error due to regression (cf. Campbell and Stanley 1966:15-16, 49-50, 70-71; Thorndike 1942).

to do the extensive cross-checking needed to establish interrater reliability for this information.) Wolf, however, argues that these data may very well be reliable because his two principal informants "are old enough to have known all of the women in the sample in their youth and are attuned to local gossip because they are often called upon to act as mediators and go-betweens" (1970:510). In any case, Wolf's conclusions do not rest on this index alone since he is able to support it with the other two kinds of information already discussed.

Wolf's measures possess a compelling amount of face validity. It is reasonable to expect that fertility, divorce, and adultery are all to some extent functions of sexual attraction, although, to be sure, each may reflect other factors as well. Our intuitive confidence in the validity of these three measures is further strengthened by the fact that all three indicators are consistent with one another, as they would be expected to be if they are all measuring the same underlying trait of sexual attraction. Once again, however, more formal evidence of the validity of these indicators as measures of sexual attraction would be highly desirable, but given the very sensitive nature of the subject, formal validation procedures would no doubt be difficult to arrange.

The replicability of the study is high. Wolf clearly describes how the various measurements were obtained and the nature of the sample he worked with. Although Wolf unfortunately does not apply statistical procedures to obtain an estimate of the role of chance, it appears from the summary data he has provided that conventional tests of significance would show that the possibility of the differences between the two groups being due to chance alone is extremely small.

Achievement in Nigeria

Wolf's research illustrates one way in which a static group comparison can be strengthened—that is, by controlling confounding extraneous variables through matching of cases. Another way of dealing with extraneous variables in a static group comparison is illustrated in research carried out in Nigeria by LeVine (1966). LeVine's study concerns achievement motivation, a motive that may be defined as a desire to succeed in terms of some objective standard of excellence. Achievement motivation appears (McClelland 1961) to be an important factor affecting rates of economic growth. As a consequence, there has been considerable interest in discovering factors that produce variation in the amount of achievement motivation possessed by different individuals and in the average amounts of achievement motivation found in different societies.

LeVine has hypothesized that achievement motivation will be affected by the kind of "status mobility system" in the society in which one lives. Since achievement motivation is essentially a deep-seated concern with being successful according to, or in terms of, some objective standard of excellence, he reasoned that cultures which, by virtue of their social organization, permitted people to rise in accordance with their ability to compete successfully with an objective standard, would also be cultures where people tended to have high achievement motivation. Accordingly, and in view of evidence from other research showing a relationship between an interest in business success and high levels of achievement motivation, LeVine hypothesized that societies in which upward mobility was achieved primarily as a result of one's ability in a pecuniary or occupational endeavor would

have higher levels of achievement motivation than would societies in which success in the political arena was stressed. LeVine's independent variable may be termed the "pecuniary-ness" of the status mobility system.⁴ The basic hypothesis may be diagrammed as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
"Pecuniary-ness" of the status mobility system in a culture	increases	achievement motivation among persons in that culture.

A measure of the dependent variable, achievement motivation, was obtained by asking informants, who were male students in secondary schools, to write essays describing dreams they had had. These dream reports were then scored for imagery indicative of strong motivation to achieve.⁵ An example of a dream description judged as containing achievement imagery is the following:

From my childhood, I have been dreaming dreams but there is one particular dream which I dream often. Probably I may dream about it this night. The dream is in fact very simple but quite funny. Whenever I dream, I see myself as Albert Onyeonwuna, the Nigerian international footballer.

I saw myself as Albert Onyeonwuna being made the captain of Nigerian team in a football-soccer competition against Ghana. I dreamt that this match was played at Lagos stadium and that the Nigerian "Red Devils," with me as captain, defeated Ghana "Iron Gates" by two goals to nil. The match was played on a rainy day and the slippery ground added fun to what could be described as the most exciting match ever played in Nigeria.

Both teams started at a fast pace and with grim determination so that it was difficult to choose between the two. Suddenly I, as Onyeonwuna, came into the picture. I dashed for the ball and after beating Ghana's right fullback, sent in a fiery shot which left Ghana's goalkeeper sprawling on the ground. The people who were in the stadium jumped up from their seats in anxiety and started clapping their hands. From that time, there were anxious moments whenever I was in possession of the ball, for the people wanted me to score more goals. The applause had hardly died down when I received a short pass from our inside right winger, Emuke. I stopped the ball, and after beating Ghana's right halfback man, parcelled the ball with a short trip to our man in the centre forward. He was in the position to kick the ball when he was kicked down by the Ghana fullback man, and this attracted the referee for a penalty kick. I was the person who played the penalty kick and it was a clean goal. Thus I raised the goal tally for our team to two. We had only two minutes to go in the match

⁴ The reasoning behind this hypothesis is complex and is related to general theory about achievement motivation. Without going into details, it may be noted that LeVine sees child training as the efficient cause of achievement and therefore the "pecuniary-ness" of the status mobility system only indirectly affects achievement by means of its impact on child-training practices. The reader interested in these issues should consult LeVine's clearly written discussion (1966). We have elected also to minimize LeVine's discussion of data collected from a third group, the Yoruba. These data support LeVine's hypothesis, but to discuss them in detail here would deflect us from the more central issue of the research design LeVine used. Some of the Yoruba data, however, are introduced later on in this discussion to illustrate other characteristics of LeVine's research design.

⁵ See LeVine 1966:103 ff. for details of the scoring system.

and Ghana players, now tired and playing short of one man, were forced into a defensive game; thus we maintained the lead of two nil until the final whistle.

This is, in fact, the dream I dream most, and whenever I wake up the following morning, I become unhappy for I realize that I am not Onyeonwuna, but Alozie.⁶

An example of a dream description judged to lack achievement imagery is this one:

There is a certain dream which I cannot sleep for a month without dreaming of it since about two years now.

I have several times seen myself flying like birds in a dream. The first day I dreamt the dream, I was going to a certain place and I was attacked by a dangerous animal. To my surprise, I flew away and the animal could not harm me, as it did not know how to fly.

On one other day I was in a dream, there was something like war. An aeroplane was flying through our town and people were shouting that the enemies had come to throw bombs. I took a kind of gun and flew away. After I had gone up to a certain height, I shot the aeroplane down and there was no effect. Then we defeated our enemies.

In fact I have always seen myself flying like birds in a dream that I do not take notice of it again because it is almost constant. Whenever in my dream I have encountered any danger, I fly away like birds from the ground to the air.⁶

LeVine was able to devise a static-group-comparison test of his hypothesis by collecting data from cultures with status mobility systems which differed traditionally in respect to the relevance of pecuniary as opposed to political success. One of the cultures LeVine chose for study was Igbo culture. The Igbo reside in eastern Nigeria and traditionally featured a mobility system where it was possible for an individual to greatly increase his status by competing with others in regard to an objective standard of success—namely, through individual achievement in commercial or other pecuniary activity. A second culture studied by LeVine was Hausa culture, where traditionally it was not generally possible for a man to achieve high status by means of individual pecuniary achievement. Instead, in Hausa culture a person could rise only through clientage—that is through a relationship of diffuse loyalty and obedience to a person of high status who was or might become a powerful officeholder in the political arena.

The static group comparison LeVine devised using groups of informants drawn from these two cultures may be diagrammed as follows:

Igbo students (exposed to a pecuniarily oriented status mobility system)	X	O
	(exposure to pecuniarily oriented status mobility system)	
Hausa students (not exposed to a pecuniarily oriented status mobility system)		O

⁶ From *Dreams and Deeds* by Robert A. LeVine (pp. 54-55). Copyright © 1966 University of Chicago Press. Used by permission.

In comparing the frequency of achievement themes in stories written by the students from the pecuniarily oriented Igbo culture to that in stories by students from the politically oriented Hausa culture, LeVine found, as predicted by his hypothesis, that the former group had more achievement themes—43 percent as compared to 17 percent. A statistical test indicated that a difference of this magnitude could be expected to occur by chance alone less than one time in one thousand. LeVine was thus able to dispose of the rival hypothesis of chance. However, at this point he still had to contend with a number of other rival hypotheses.

Due to the nature of the research situation, the treatment and comparison groups, which were drawn from different cultures, were unavoidably affected in different ways by a number of extraneous variables, some of which produced plausible rival hypotheses. One such rival hypothesis was that the difference in achievement motivation between the groups might have been due to a religious factor. The Igbo treatment group was overwhelmingly Protestant Christian in religion, while the Hausa comparison group was predominantly Moslem. It might conceivably be maintained that Protestant Christianity is a religion which encourages individual achievement, while Islam lays more stress on obedience, thereby discouraging individual achievement, and that this difference in the two groups' religious orientation suffices to explain their differences in achievement motivation.

LeVine dealt with this rival hypothesis in the following manner. He had, as part of his study, collected information on a group of students from a third culture in Nigeria—the Northern Yoruba. It happens that some Northern Yoruba are Christian and some are Moslem. LeVine reasoned that if exposure to Protestantism, as opposed to Islam, was a factor in determining level of achievement motivation, it should be possible to demonstrate this by means of a static group comparison of Christian Northern Yoruba and Moslem Northern Yoruba. This comparison may be diagrammed as follows:

Christian Northern Yoruba	X	O
	(exposure to Protestant Christianity)	
Moslem Northern Yoruba		O

The results of this static group comparison revealed absolutely no indication that Protestant Christianity promoted achievement motivation any more than did Islam (Table 3.5). LeVine inferred from this result that it was improbable that the

TABLE 3.5 PERCENTAGE OF AI^a DREAMERS AMONG THE NORTHERN YORUBA ACCORDING TO RELIGIOUS AFFILIATION^b

	<i>Moslem</i>	<i>Christian</i>	<i>Totals</i>
AI	27.2(3)	27.2(6)	9
Non-AI	72.8(8)	72.8(16)	24
Totals	100 (11)	100 (22)	33

^a AI = Achievement Imagery.

^b Adapted from LeVine, 1966:59, Table 5. Copyright © 1966, University of Chicago Press and used with permission.

difference in achievement motivation between the Igbo and the Hausa could be the result of the extraneous factor of religious differences. He therefore rejected this rival hypothesis. (But see p. 47 for a cautionary note.)

A second rival hypothesis was related to the fact that the parents of the Igbo informants had generally received much more Western education than had those of the Hausa. It could be hypothesized that this extraneous factor was capable of accounting for the difference in achievement motivation between the two groups. A rationale for such a hypothesis might be that Western education stresses individual achievement. Therefore, if a child's parents were Western educated, it is likely they would give more stress to achievement in their training of the child.

LeVine was able to discredit this rival hypothesis in a manner analogous to his procedure regarding the rival hypothesis involving religion. He found that there was enough variability in terms of parental education within the group of Igbo informants to permit a static-group-comparison test of the hypothesis that having Western-educated parents would increase achievement motivation. Igbo boys who had mothers⁷ with little or no Western education were compared with those whose mothers had substantial Western education; that is:

Igbo with mothers having substantial Western education	X	O (exposure to a mother with substantial Western education)
Igbo with mothers having little or no Western education		O

This static group comparison showed that having a Western-educated mother had no appreciable effect (Table 3.6). LeVine therefore dismissed as improbable the rival hypothesis of parental education.

Finally, by following a procedure analogous to that described for the two rival hypotheses treated above, LeVine also was able to reject as improbable a third rival hypothesis which involved special drive-arousing circumstances affecting the Hausa but not the Igbo.

The technique of discounting rival hypotheses used by LeVine—testing the rival hypothesis in a different situation and showing it to be unsupported in that situation—can be referred to as “control by testing.” Control by testing is generally less powerful than control by matching,⁸ which we discussed in connection with Wolf's study. This is the case because showing that a factor has no effect in one situation

⁷ LeVine compared the boys in terms of their mothers' education rather than their fathers' education or fathers' and mothers' combined education on the grounds that the mother's education was “. . . theoretically the most powerful influence on personality development, since the mother has the most contact with the child . . .” (1966:57).

⁸ We use “control by matching” to refer both to the actual matching of cases and to the statistical counterparts of matching such as covariance analysis and partial correlation.

TABLE 3.6 MOTHERS' EDUCATION AND PERCENTAGE OF SONS REPORTING ACHIEVEMENT IMAGERY DREAMS^a

	Percent reporting dreams with achievement imagery	Percent reporting dreams without achievement imagery
Igbo with mothers having up through five years of Western education	45 (46) ^b	55 (56)
Igbo with mothers having more than five years of Western education	43 (15)	57 (20)

^a Adapted from LeVine 1966:58, Table 4. Copyright © 1966 University of Chicago Press. Used by permission.

^b Numbers in parentheses are numbers of informants.

does not guarantee it will have no effect in a different situation. Thus, showing that Protestant Christianity as opposed to Islam has no effect on achievement among the Northern Yoruba does not guarantee religion did not cause a difference in achievement motivation between the Igbo and the Hausa. Similarly, showing that mother's education does not influence achievement motivation within the Igbo group does not guarantee that this factor does not contribute to the difference in achievement motivation between the Igbo and the Hausa. Demonstrations such as these, however, do *reduce* the likelihood that the extraneous factors involved are important determinants of the dependent variable. Thus, control by testing is well worth applying when control by matching is impossible.

LeVine also identified but was unable to positively reject several other rival hypotheses deriving from extraneous variables that may have affected the two groups in different ways. One of these concerned population pressure. The Igbo are subject to extreme population pressure on limited land resources. The Hausa experience substantially less population pressure. It might be argued that the high population pressure in Igbo-occupied areas and the attendant threat of poverty could give rise to a high level of motivation to achieve, especially in regard to commercial and other nonagricultural activities. LeVine argues against this and two additional rival hypotheses, but is unable to dismiss them since neither control by matching nor control by testing could be applied to the extraneous factors involved, given the limitations of the particular research situation. Thus, these hypotheses remain as untested but reasonable alternatives to the status mobility system hypothesis advocated by LeVine. Though such results might seem at first to be cause for discouragement, they are for LeVine (as for others familiar with the nature of scientific inquiry) a stimulus for research. LeVine concludes his study by urging the design of further research that will take into account these alternative hypotheses.

Selection also posed a problem in LeVine's study, since the several schools from which LeVine's student informants were drawn varied considerably in their entrance requirements. The entrance requirements for the government colleges (which, as LeVine suggests, “are universally regarded as the schools with the best students”) are lower in the northern part of Nigeria where the Hausa are concentrated than

they are elsewhere in the country. Thus, it could be argued that the Hausa students are "likely to be of lower caliber than their Ibo [sic] counterparts" (1966:45). If this is true, and if "lower caliber" implies (among other things) a lower degree of achievement motivation, selection becomes a plausible rival hypothesis explaining the difference between the Hausa and Igbo. LeVine suggests that selection bias due to the government colleges' differing entrance requirements may be offset by other circumstances relating to selection of the sample. And so, while he is not able to rule out selection as a rival hypothesis, he discounts it on the grounds that when these other factors are considered, "there is no convincing evidence that the selection of subjects biased the results . . . (1966:46)."

We can assess LeVine's study in terms of our general evaluative criteria as follows. The paradigm chosen for the study, the static group comparison, appears to be the most powerful that is feasible, given the nature of the research problem. A more powerful paradigm such as the nonequivalent-control-group paradigm or the standard-control-group paradigm would require premeasurement of the dependent variable. But this is unfeasible since, as was the case in Wolf's study, the independent variable begins to act in early childhood, if not in infancy. The study attains a high level of replicability since all stages of the research and the various procedures employed are explained in detail (for example, LeVine includes an appendix describing the scoring of dream reports for achievement motivation). Both intrarater and interrater reliabilities are provided for the scorings of the dream reports, and they prove to be satisfactorily high. There is no formal demonstration of validity for the dream measure of achievement motivation. However, the measure was closely patterned after another technique for measuring achievement motivation that has been shown to be valid (see McClelland et al. 1953). The dream descriptions also were scored by raters who knew nothing of the hypothesis of the study, eliminating another potentially invalidating factor. The role of chance was taken into account by application of appropriate statistical tests, and this procedure indicated that the rival hypothesis of chance could be conclusively rejected.

To conclude our discussion of LeVine's study we may note that LeVine has been able to strengthen greatly the case for his hypothesis relating type of status mobility system to achievement motivation, although he has *not* been able to reject *all* rival hypotheses. While some plausible rival hypotheses remain unrefuted by this study, it is to be hoped that future research in different settings will be successful in testing them. If this is done and the rival hypotheses that LeVine was unable to deal with are shown to be invalid, the case for LeVine's hypothesis will be still stronger.

Child-training and Food Accumulation

Many hypotheses of interest to anthropologists cannot be tested within a single culture because they deal with factors that do not vary sufficiently within such relatively limited bounds. For example, hypotheses concerning type of descent system can seldom be tested within the confines of a single culture since cultures tend to be homogeneous with regard to this variable—that is, all (or virtually all) Chinese reckon kinship patrilineally, all Trobrianders reckon kinship matrilineally,

and so on. Testing hypotheses involving such variables generally requires a cross-cultural approach—in other words, information must be gathered from more than one culture. LeVine's study, which we have just discussed, provides one example of the cross-cultural approach. Since LeVine wanted to test a hypothesis concerning the effects of a culture's status mobility system on individuals' achievement motivation, and since a status mobility system is more or less uniform within a particular culture, he found it necessary to use informants drawn from two cultures in order to obtain treatment and comparison groups that were exposed to sufficiently differing degrees of the independent variable. The cases in LeVine's research were still individuals, even though the two groups of individuals being compared were drawn from different cultures.

In many other cross-cultural studies the cases are not individuals but cultures.⁹ A good example of this latter kind of cross-cultural study is provided by Barry, Child, and Bacon (1959). This study is concerned with the relationship between accumulation of food resources and child-training. The authors define food accumulation as "the degree to which the food resources . . . are characteristically present in advance to be cared for or stored prior to being used, as against being consumed as soon as procured" (1959:53). They view the degree of food accumulation in a culture as a function of the type of subsistence economy it features. Cultures where subsistence is based on animal husbandry are assumed to be high in food accumulation since, in such cultures, "the meat that will be eaten in coming months and years, and the animals that will produce the future milk, are present on the hoof" (1959:52). Agricultural cultures are assumed to be moderately high on food accumulation since the harvest must be stored for future consumption. Cultures where the subsistence economy is based on hunting or fishing and that possess no means for preserving the catch are assumed to have a low degree of food accumulation.

Barry and his coauthors reason that in high-food-accumulation cultures, adults would tend to conform strongly to traditional routines designed to preserve the food supply and that innovation regarding economic activities would be discouraged. This would occur in such cultures, they argue, because ". . . carelessness in performance of routine duties leads to a threat of hunger, not for the day of carelessness itself but for many months to come" (1959:52).

In low-food-accumulation cultures, conversely, there is less likelihood of this danger arising from venturesomeness and initiative. If a hunter or fisherman tries a new approach, he may, if it fails, lose his day's catch. He can, however, immediately revert to traditional techniques in order to get the next day's catch. On the other hand, if the innovation is successful and produces a better catch than the traditional techniques, the hunter or fisherman is immediately benefited. Such cultures should, therefore, place a premium on innovativeness and assertion.

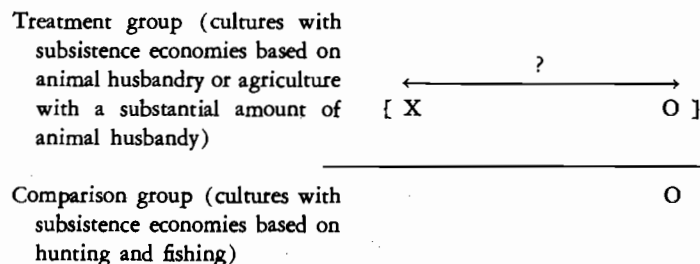
⁹It should also be noted in conjunction with our discussion of cross-cultural studies which treat entire cultures as cases, that determining what constitutes a single "culture" can prove highly problematic. A detailed discussion of the problem of identifying basic culture units (which is part of the so-called Galton-Flower's problem) is beyond the scope of this work. The interested reader is referred to the specialized treatments in Rohner (n.d.) and Naroll and Cohen (1970).

On this basis, Barry and his colleagues developed the hypothesis that high-food-accumulation cultures will feature child-training practices which emphasize compliance, while low-food-accumulation cultures will emphasize assertiveness in their child training. Following our usual format, we may state the hypothesis as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Accumulation of food	increases	stress on compliance as opposed to assertiveness in child-rearing practices.

In order to test this hypothesis, cultures that had subsistence economies based on animal husbandry or on agriculture combined with a substantial amount of animal husbandry were classed as "extremely high" in food accumulation. Hunting and fishing cultures were classed as "extremely low" in food accumulation. Measurement of the dependent variable involved a systematic assessment, made by two independent raters, of the relative preponderance of compliance or assertion in child-training practices. The raters did not know of the hypothesis to be tested at the time they assessed the ethnographic literature, although the degree of interrater reliability is, unfortunately, not reported. On the basis of these ratings, each culture in the sample was given a score on the dependent variable of compliance versus assertion. The scores ranged from -15 to +13.5. A high positive score indicates a high degree of compliance, while a high negative value indicates a high degree of assertiveness.

We can diagram Barry, Child, and Bacon's study as follows:



The question-marked double-headed arrow and the brackets around the X and O for the treatment group are used here to indicate that, *strictly speaking, we are only assuming the independent variable preceded the dependent variable*. Due to the lack of historical information, we cannot know for certain that a change in a system of high food accumulation occurred first and was followed by stress on compliance in child training in the various cultures that make up the treatment group. It is conceivable the sequence was actually the reverse. Thus, it might be argued that child training which stressed compliance developed first and caused the affected

cultures to acquire subsistence economies that feature a high degree of food accumulation. This latter sequence, while conceivable, is sufficiently implausible that no one, to our knowledge, has seriously proposed it.

The results of the static group comparison showed, as predicted in the hypothesis, that the cultures in the treatment group, the group of cultures high in food accumulation, evidenced a much higher emphasis on compliance in their child training. This is graphically illustrated by Table 3.7 (p. 52). The number following each culture is its rating on the scale of compliance versus assertiveness. As may be ascertained from the table, the average score for the treatment group was 5.25 while that for the comparison group was -4.82. Statistical analysis indicated that the rival hypothesis of chance could be rejected since there was less than one chance out of a thousand that a difference this great would occur by chance alone.

It is sometimes argued that studies such as Barry, Child, and Bacon's, which use individual cultures as cases, largely avoid the problem of extraneous variables since these variables are "randomized out." Those who take this position would hold that since both the treatment and comparison groups are composed of many different cultures, it is difficult to conceive of any extraneous variables affecting one group of cultures substantially more than the other. Since both groups can be expected to be affected in roughly similar fashion by extraneous factors, so the argument goes, the possibility of error due to extraneous variables is remote.

There may be a grain of truth in this argument. The probability of error due to extraneous factors is no doubt somewhat less in such a study than in one like LeVine's, where the treatment and comparison groups may be expected to differ systematically on a great number of variables because all the cases in a given group are drawn from one culture. However, it must not be assumed that the problem of extraneous variables is totally eliminated when numerous cultures are used to supply the cases for a given group. It is still possible for treatment and comparison groups to be differentially affected by extraneous variables even though each group is composed of many different cultures. Thus, in the present example, Barry and his coauthors provide evidence that the cultures in the treatment group, in addition to being higher in food accumulation relative to the comparison group, also tend to be higher on variables such as degree of political integration, complexity of social stratification, and size of settlements. It is likely that the treatment and comparison groups differ on other extraneous variables as well—such as settlement density. Plausible rival hypotheses involving such extraneous factors might be offered as alternatives to the food-accumulation hypothesis. For example, such a hypothesis might be: In societies that have a high degree of settlement density, children will be trained to be nonassertive in order to minimize fighting between children of different families, which might lead to dangerous adult conflict.¹⁰

In evaluating this study according to our general criteria, we may note the following points. The paradigm employed appears to be the strongest possible under

¹⁰ The researcher may, of course, attempt to deal with such rival hypotheses deriving from extraneous factors by control through matching or control through testing. While Barry et al. have not, strictly speaking, used either of these techniques, they have attempted, through statistical analysis, to demonstrate that rival hypotheses involving a number of extraneous factors are implausible (*cf.* Barry et al. 1959:59 f.).

TABLE 3.7 RELATION OF SUBSISTENCE ECONOMY TO GENERAL PRESSURE TOWARD COMPLIANCE VERSUS ASSERTION IN CHILD-TRAINING^a

Extremely high accumulation cultures	Extremely low accumulation cultures
Aymara (+13½)	
Tepoztlan (+13½)	
Lepcha (+11½)	
Swazi (+ 8½)	
Tswana (+ 8½)	
Nyakyusa (+ 8)	
Sotho (+ 8)	
Nuer (+ 7)	
Tallensi (+ 7)	
Lovedu (+ 6½)	
Mbundu (+ 6½)	
Venda (+ 6½)	
Kikuyu (+ 6)	
Zulu (+ 6)	
Pondo (+ 4½)	
Chagga (+ 4)	
Ganda (+ 3)	
Chamorro (+ 2½)	Teton (+ 4)
Masai (+ 2½)	Yahgan (+ 1)
Chukchee (+ 1)	Hupa (+ ½)

Median ^b	
Tanala (0)	Chiricahua (0)
Thonga (- 2½)	Murging (0)
Araucanian (- 3)	Paiute (0)
Balinese (- 3)	Arapaho (- 2)
	Kwakiutl (- 2)
	Cheyenne (- 2½)
	Kaska (- 2½)
	Klamath (- 2½)
	Ojibwa (- 2½)
	Ona (- 3)
	Aleut (- 4)
	Jicarilla (- 6½)
	Western Apache (-10)
	Siriono (-10½)
	West Greenland
	Eskimo (-11)
	Aranda (-12)
	Comanche (-12)
	Crow (-13½)
	Manus (-15)

^a The cultures are "listed within each column in descending order of degree of pressure toward compliance as compared with pressure toward assertion. The number in parentheses after each society indicates the degree of preponderance of compliance (plus scores) or of assertion (minus scores)" (Barry et al. 1959:60).

^b The dashed line marks the median point. When a set of scores is arranged in order according to the value of each score, the point that divides the ordering into two equal groups (one half considered to be "high" scores and the other "low" scores) is known as the *median*.

Adapted from Barry et al. 1959:60, Table 2. Reproduced by permission of the American Anthropological Association from the *American Anthropologist*, 61(1), 1959.

the circumstances. The replicability of the study is generally good, since details of sampling and measurement procedures are supplied. However, the reliability of the measure of assertion-compliance in child-training was not reported, although inter-rater reliability, at least, could have been determined since two raters were used in the study. Similarly, no formal evidence for the validity of the measure is provided. Finally, the role of chance was taken into account by means of appropriate statistical procedures.

Correlational Analysis as a Variant of the Static-Group-Comparison Paradigm: Achievement Motivation and Rates of Economic Growth

An approach to hypothesis-testing commonly employed in anthropological research is correlational analysis. Correlational analysis, in the form in which it is most commonly applied, can be viewed as a variant of the static-group-comparison paradigm. The major difference is that in correlational analysis many levels of the independent variable are utilized instead of only two as in the standard static group comparison.

To exemplify this difference, let us consider research by McClelland (1961) on achievement motivation and economic development. McClelland proposed the hypothesis that the rate of economic development of a country would be influenced by the level of achievement motivation among its people. This hypothesis may be stated as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
The higher the level of achievement moti- vation among the people of a country	the higher its	rate of growth.

Achievement motivation, as you will recall from our discussion of LeVine's research in Nigeria, has been defined by McClelland as the degree to which a person desires to succeed in terms of some objective standard of excellence.

McClelland was able to measure achievement motivation by analyzing themes in children's readers. Each country in the study was assigned a score on achievement motivation on the basis of the frequency of achievement themes and achievement imagery in its children's readers in 1925. (The assumption was that the frequency of achievement themes and related imagery in a country's schoolbooks was an index of the average level of achievement motivation of the people in that country.) He derived a measure of economic development from the increase in the amount of electricity used per capita in each country over the following twenty-five year period. (The assumption behind this measure was that the amount of electricity used per person was an indication of how much modern machinery was in use and therefore an indicator of the degree of economic development in that country.) McClelland's measurements for twenty-two countries are reproduced in Table 3.8.

TABLE 3.8 ACHIEVEMENT MOTIVATION AND ECONOMIC DEVELOPMENT^a

Country	Scores on the independent variable, achievement motivation level, 1925	Scores on the dependent variable, economic development, 1950
Ireland	3.19	0.33
Australia	2.81	1.13
Canada	2.67	1.73
Cases above Sweden	2.19	3.17
Median on Great Britain	2.10	1.65
Achievement Motive Denmark	2.00	0.14
U.S.A.	1.90	1.86
Argentina	1.86	-0.61
Austria	1.57	-0.12
New Zealand	1.48	1.86
Uruguay	1.48	-0.62
Median		
Germany	1.38	-0.79
Norway	1.33	-0.03
Hungary	1.29	-0.26
Cases below Chile	1.29	-0.43
Median on Finland	1.24	0.74
Achievement Motive Union of S. Africa	1.05	0.69
Belgium	1.00	-0.75
France	0.81	-0.55
Spain	0.81	-0.63
Greece	0.38	-0.52
Netherlands	0.29	-0.10

^a Cases are arranged according to their achievement motivation scores.

Adapted from McClelland, 1961:90-91, Table 3.4. From *The Achieving Society* by D. McClelland © 1961 by Litton Educational Publishing, Inc. Reprinted by permission of Van Nostrand Reinhold Company.

Using McClelland's data it is possible to construct a standard static group comparison. We can distinguish a treatment group, defined as those countries scoring high (that is, above the median) on the independent variable of achievement motivation and a comparison group, defined as those countries scoring low (below the median) on this variable. When this is done, we have, in effect, a standard static group comparison:

Treatment group (the 11 countries above the median on achievement motivation)	X	O
	(a high level of achievement motivation)	
Comparison group (the 11 countries below the median on achievement motivation)		O

If we calculate the average economic development scores for the two groups we find that, consistent with the hypothesis, the treatment group scores more than twice as high on economic development (2.11) as the comparison group (0.99).

Casting these data into the form of an ordinary static group comparison is not, however, the most effective form of analysis. When quantitative or rank-order measurement of the independent variable has been achieved, as in this case, collapsing the measurements into two categories—"high" and "low"—obviously discards a great deal of information. That is, we know some of the cases in the "high" group have much more of the independent variable than others, yet we ignore this information and treat all the "high" group cases as having the same amount of this variable. Correlational analysis makes it possible to test for the hypothesized relationship using *all* the available information.

As we noted, correlational analysis, as it is usually applied, can be viewed as a type of static group comparison that uses many levels of the independent variable instead of only two as in the standard static group comparison. When we say that the standard static group comparison uses only two levels of treatment, we refer to the fact that one group, the treatment group, is exposed to a *high* level of the independent variable and another group, the comparison group, is exposed to a *low* level of the independent variable. By contrast, in correlational analysis, we deal with many levels of the independent variable. These two approaches to hypothesis-testing can be compared in the diagrams on page 56.

In the diagram for the correlational approach each of the X's represents a different level of the independent variable. The actual number of levels in a particular correlational analysis will depend upon how many different scores there are for a particular independent variable. If there are ten different scores on the independent variable in a sample, the correlational analysis will have ten different levels. Two or more cases may, of course, have the same score on the independent variable, in which case they will fall at the same level.

Now let us apply correlational analysis to the present example of McClelland's research. Instead of dividing the cases into two groups to construct a standard static group comparison, as before, we will this time allow each case to retain its individual score on the independent variable (as McClelland did). Of the 22 countries in McClelland's sample, three pairs of countries (New Zealand-Uruguay, Hungary-Chile, France-Spain) are tied on the independent variable. Thus, only 19 levels of the independent variable (achievement motivation) are represented in this sample of countries.

McClelland's hypothesis predicts that the dependent variable, economic growth, will increase as the independent variable of achievement motivation increases. Therefore, when we apply the correlational approach to these data, we would expect that, if the hypothesis is correct, we will observe an increase in the dependent variable as we go from the lower levels of the independent variable to the higher levels. In other words, there should be a tendency for cases high on the independent variable to be high on the dependent variable and for cases low on the independent variable to be low on the dependent variable.

We can form some idea of whether such a tendency exists by visually examining

STANDARD STATIC GROUP
COMPARISON

(Uses only two levels of the independent variable, "high" and "low.")

Treatment Group (exposed to a <i>high</i> level of the inde- pendent variable)	X	O
Comparison Group (exposed to a <i>low</i> level of the inde- pendent variable)		O

CORRELATIONAL APPROACH

(Uses many levels of the independent variable. Here each level is represented by a separate subscripted X.)

X _{level n}	O
X _{level 6}	O
X _{level 5}	O
X _{level 4}	O
X _{level 3}	O
X _{level 2}	O
X _{level 1}	O

a table of scores. Examination of Table 3.9 shows that there does appear to be a tendency for countries high on achievement motivation to be high on economic development and vice versa, although there are some nonconforming cases—for example, Argentina. As you can see, it is difficult to judge the strength of such a tendency by visual inspection alone. What is needed is some concise index of the strength of the tendency for two variables to vary together—that is, an index of the degree to which they are correlated (*co-related*). Mathematical procedures have been developed as an aid to correlational analysis and provide several indices of the strength of the relationship between two variables. These indices are known as correlation coefficients. One of the most widely used¹¹ correlation coefficients is Pearson's *r*. Pearson's *r* may range in value from -1.0 to $+1.0$. An *r* that approaches $+1.0$ indicates a very strong tendency for one variable to *increase* as the other increases; a value near 0.0 indicates there is almost no tendency for one variable to increase as the other increases; and a value near -1.0 indicates a very strong tendency for one variable to *decrease* as the other increases.

When a Pearson's *r* correlation coefficient is computed for the McClelland data it yields a value of $.54$, indicating a moderately strong tendency for the dependent variable to increase as the independent variable increases. However, in correlational analysis, as in other research, we must be alert to the role of chance. It is possible for chance alone to produce the appearance of some correlation between dependent and independent variables in a sample of cases when in fact no rela-

¹¹ Pearson's *r* has won such wide acceptance as a measure of correlation that in the social science literature, whenever reference is made to the "correlation coefficient" this is understood as a reference to Pearson's *r*. Other measures of correlation have been developed, however.

TABLE 3.9 ACHIEVEMENT MOTIVATION AND ECONOMIC DEVELOPMENT^a

Level of independent variable	Country	Scores on the independent variable, achievement motivation level, 1925	Scores on the dependent variable, economic development, 1950
Level 19	Ireland	3.19	0.33
Level 18	Australia	2.81	1.13
Level 17	Canada	2.67	1.73
Level 16	Sweden	2.19	3.17
Level 15	Great Britain	2.10	1.65
Level 14	Denmark	2.00	0.14
Level 13	United States	1.90	1.86
Level 12	Argentina	1.86	-0.61
Level 11	Austria	1.57	-0.12
Level 10	New Zealand	1.48	1.86
	Uruguay	1.48	-0.62
Level 9	Germany	1.38	-0.79
Level 8	Norway	1.33	-0.03
Level 7	Hungary	1.29	-0.26
	Chile	1.29	-0.43
Level 6	Finland	1.24	0.74
Level 5	Union of S. Africa	1.05	0.69
Level 4	Belgium	1.00	-0.75
Level 3	France	0.81	-0.55
	Spain	0.81	-0.63
Level 2	Greece	0.38	-0.52
Level 1	Netherlands	0.29	-0.10

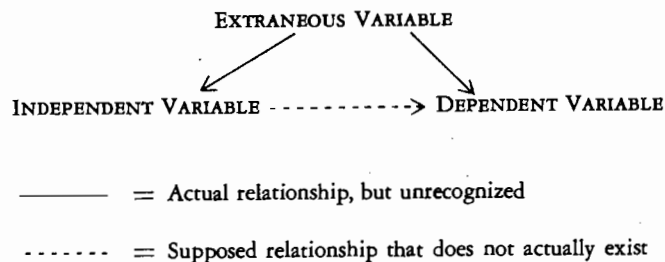
^a Cases are arranged according to their achievement motive scores and the different levels of the independent variable are distinguished. Adapted from McClelland, 1961:90-91, Table 3.4. From *The Achieving Society* by D. McClelland © 1961 by Litton Educational Publishing, Inc. Reprinted by permission of Van Nostrand Reinhold Company.

tionship exists. Statistical procedures have been developed to estimate the role of chance in correlational analysis (Hardyck and Petrinovich 1969:203 f.). By applying these we find that so strong a correlation between the two variables in the McClelland study would be obtained by chance alone less than one time out of a hundred. Therefore, we are probably safe in rejecting chance as a factor here.

Thus, this correlational analysis, which makes full use of the information contained in the data, supports the hypothesis that a high level of achievement motivation produces a high rate of economic growth. However, the support for this hypothesis, like that found for any other, is subject to the limitations of the structure of the research. Let us consider, therefore, the types of error to which a correlational analysis of the type discussed here is particularly vulnerable.

Although a correlational analysis may be viewed as a variant of the static group comparison, two kinds of error that plague the standard version of this paradigm—selection and interaction effects involving selection—do not apply to correla-

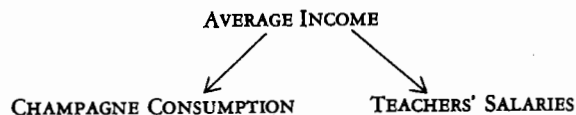
rional analysis since there are no treatment and comparison groups. All the cases in the typical correlational analysis are considered collectively. However, correlational analysis is *highly subject to errors resulting from extraneous variables*. It is this potential source of error that lies behind the oft-repeated refrain, "correlation doesn't prove causation." That is, a high correlation between an independent and dependent variable may be entirely due to the effect of some extraneous variable that varies together with the independent variable. We might, for example, encounter a situation in which the independent variable has no effect on the dependent variable, but both the independent and dependent variables are affected by a third extraneous variable:



This would produce a high correlation between the independent and dependent variables since they are simultaneously and similarly influenced by the extraneous variable. Because of this, a researcher would be in danger of mistakenly concluding that the independent variable affects the dependent variable. An example of such a situation would be a high correlation between champagne consumption and teachers' salaries. That is, it might be that in cities where champagne consumption is high, teachers' salaries also tend to be high, and vice versa. A rash investigator might be tempted to view champagne consumption as an independent variable affecting teachers' salaries, on the grounds that school board members, when happily inebriated, will tend to grant higher salaries to teachers:

CHAMPAGNE CONSUMPTION -----> TEACHERS' SALARIES

However, such a correlation would almost certainly result from an extraneous factor of average income, which affects both champagne consumption and teachers' salaries:



This kind of error in correlational analysis is analogous to the error that may arise in a standard static group comparison when the treatment and comparison groups are differentially affected by extraneous variables (when there is selection-extraneous-variable interaction).

The effects of particular extraneous variables in a correlation analysis may be eliminated by a special procedure known as partial correlation. In this technique, the effects of the extraneous variable are removed mathematically. This technique thus fills somewhat the same role in correlational analysis as does matching in the standard static group comparison, making it possible, in effect, to control specific extraneous variables.

A further difficulty often encountered in correlational analysis is uncertainty concerning the direction of influence. In McClelland's study this was *not* a problem, as achievement motivation was measured in 1925, while rate of economic development was measured over the period 1925-1950. If there is a causal relationship between the two variables, then clearly the temporally earlier variable of achievement motivation must be causing the temporally later variable of economic development. It could hardly be maintained that subsequent economic development caused the level of achievement motivation in 1925. Often, however, the time sequence of the variables in a correlational analysis is not known and can only be assumed. This, of course, introduces further uncertainty into the analysis. This problem, it should be pointed out, is not unique to correlational analysis. Indeed, one of our examples of the standard static group comparison, the Barry, Child, and Bacon study, featured this type of uncertainty concerning the temporal sequence of events.

In evaluating McClelland's research according to our general evaluative criteria, we can make the following observations. The paradigm employed, the correlational variant of the static group comparison, appears to be the strongest possible given the nature of the research problem. Since it makes full use of the detailed information available on the independent variable, it is notably superior to the standard static group comparison. It is also important to note, however, that in the test of this particular hypothesis, McClelland was able to guard against only a few of the potential extraneous variables (such as the effect of World War II damage on some countries in his sample) and his findings are weakened by this.¹²

The measure of achievement motivation, the independent variable, appears to possess satisfactory reliability. The reliability of this measure was assessed in two ways. First, interrater reliability was computed. The measure involves, as we noted, scoring stories from children's readers for indications of a concern with achievement. As two judges working independently were used to score the stories, it was

¹² In fairness, however, it must also be pointed out that this hypothesis is only one of a complex array of hypotheses concerning achievement motivation that McClelland has tested. Taken together, these hypotheses make it difficult to identify particular extraneous variables that could be producing the correlations he has found. We will discuss the importance of this strategy of multiple hypothesis-testing, or the strategy of "cumulative, complementary research" in more detail later (page 105). For a recent, outstanding example of such cumulative, complementary research by anthropologists, see Cole et al. (1971).

possible to compute interrater reliability, which was found to be quite high. Second, split-half reliability was determined. The twenty-one stories for each country were divided into sets of ten and eleven stories apiece and the scores for one set were correlated with the scores on the other set. This operation revealed an acceptably high split-half reliability.

The measure of economic development, the dependent variable, also seems likely to possess a satisfactory level of reliability, since procedures for measuring electricity consumption are highly developed technically and are well standardized throughout the world.

As regards validity, the measure of achievement motivation used is a direct extension of another measure that has been shown rather convincingly to be valid (see McClelland et al. 1953). Thus, it might reasonably be maintained that the modified measure used in this study is also likely to be valid. More direct evidence of its validity is, of course, to be desired. The measure of economic development, derived from electricity consumption per capita, possesses a great deal of face validity, since electricity consumption is intimately connected with modern economic growth. The validity of this measure is further supported by the fact that it correlates quite highly with a measure of per capita real income, Colin Clark's "international unit" income measure (McClelland 1961:80-87).

The replicability of the study appears to be satisfactory. Detailed descriptions are provided for all major procedures utilized. The steps taken to deal with the rival hypothesis of chance also appear satisfactory. As we noted, a statistical test indicated that there was only a very small possibility that the results of the study were due to chance.

In concluding this review of McClelland's study, one final comment seems appropriate. It is apparent that McClelland's hypothesis does not fully account for the variation in the measure of economic growth for the sample countries, since the data show only a moderately strong relationship between the independent variable of achievement motivation and economic growth. This suggests that other factors besides achievement motivation also contribute to economic development. Many of the phenomena studied by anthropologists have multiple causes and it is therefore exceedingly rare for a simple one-factor hypothesis such as the one reviewed here to account completely for the variation in the dependent variable. Of course, it would be naive in the extreme to reject a hypothesis because it only accounted for a part of the variation in the dependent variable. What is called for is the elaboration of the hypothesis to incorporate other important independent variables as well.

THE PRETEST-POSTTEST PARADIGM

Social Contact and Attitude Change

An example of the pretest-posttest paradigm is provided by the research of Festinger and Kelley (1951). The setting for this research was a large government-

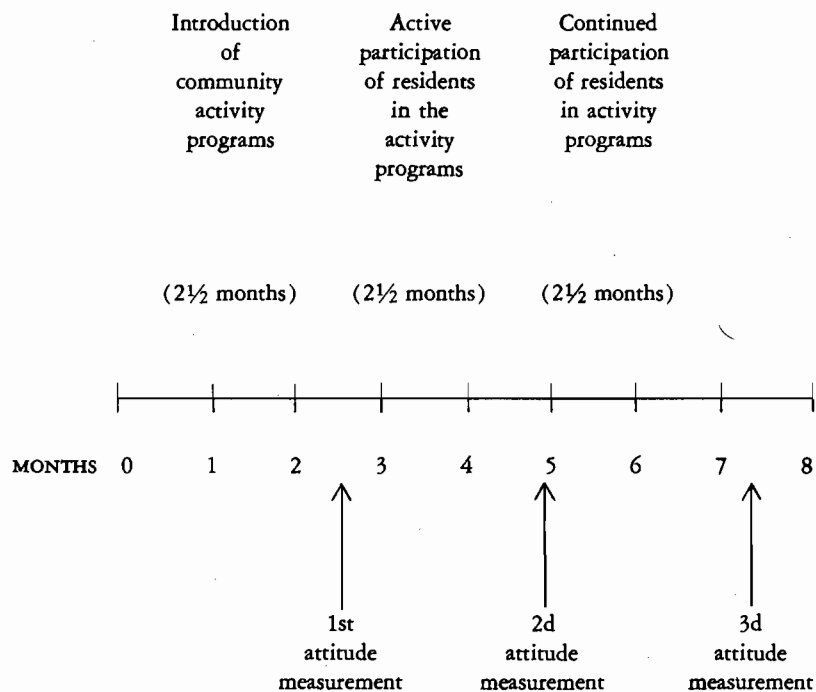
financed housing project in a small eastern American city. A research team that did a preliminary survey of the project found a low degree of social contact, with few residents enjoying active friendships within the project. Project residents, in fact, expressed a high degree of hostility toward one another, viewing their project neighbors as "low class" and "undesirable." The researchers found that project residents, as a group, were actually no lower in status than other people in the community, according to objective indicators such as education and occupation. They ascertained, however, that, originally, most project residents had not wanted to move into the project and saw themselves as having been forced to move there by various circumstances beyond their control, such as a shortage of housing. The researchers speculated that the residents' generally unfavorable initial expectations about the project had prevented them from forming many social contacts after moving into it. This lack of contact among project dwellers then perpetuated their unfavorable attitudes toward one another. Their analysis of the situation led the researchers to form the hypothesis that "an action program which stimulated contact among project residents under favorable conditions would act to break down their hostile attitudes toward each other," or, in diagrammatic form:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Contact under favorable conditions	decreases	hostile attitudes.

To test this hypothesis, Festinger and Kelley utilized a pretest-posttest paradigm. Professional community workers set up a series of community activity programs in the project, including a nursery school program and various recreational activities programs for school age children, teen-agers, and adults. The overall program of community activities began with a two and one-half month period in which the community workers explained the various activity programs to the project residents and attempted to interest them in taking part. At the end of this introductory phase, an attitude survey was carried out among a sample of project residents. Unfavorable attitudes toward other members of the project were measured by ascertaining whether persons interviewed made "hostile" comments (as judged by the researchers on the basis of responses to a standard list of interview questions) regarding other project residents. Then followed a second period of two and one-half months during which residents became actively involved in the various community programs. At the conclusion of this second period the attitude survey was repeated using another sample of project residents. The activity program then was continued for still another two and one-half month period, after which the attitude survey was administered for the third and last time.

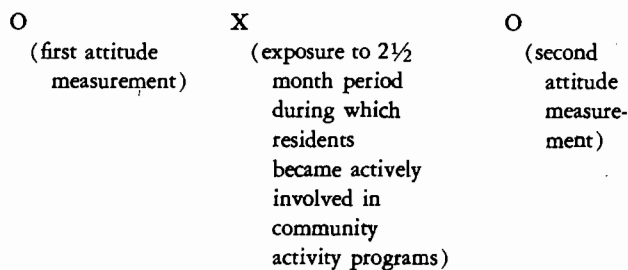
The timing of the study is summarized in Figure 3.1.

FIGURE 3.1

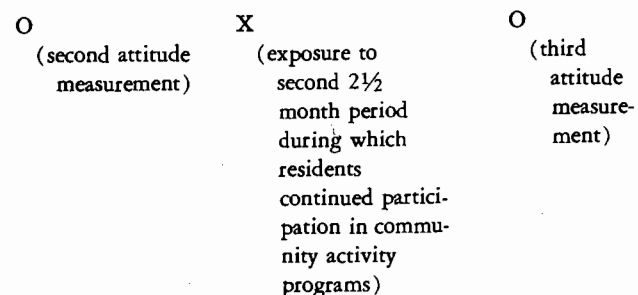


You will note that *three* measurements were carried out instead of the two that are standard for the pretest–posttest paradigm. We may view Festinger and Kelley's study as constituting two sequential pretest–posttest studies, with the middle measurement constituting the posttest for the first study and the pretest for the second. This may be diagrammed as follows:

a) First pretest–posttest study



b) Second pretest–posttest study



The results of the two pretest–posttest studies are shown in Table 3.10.

To sum up these results, the total percentage of "hostile" comments increased slightly during the first study (that is, between the first and second surveys), and increased somewhat more during the second study (between the second and third surveys). These results obviously do not support the hypothesis. If anything, they suggest that the community activities may have *increased* negative attitudes among the residents. The explanation for the failure of the hypothesis, however, does not lie in failure of the community activities programs to actually stimulate increased contact among residents. Overall, the amount of social contact was substantially increased. Survey data showed that the average number of different project residents who were invited into the respondents' homes increased from 3.7, at the time of the first survey, to 5.5 and 6.4, for the second and third surveys. The research results, then, cast doubt on the validity of the hypothesis that increased contact, per se, will reduce negative attitudes. Rather, it may be, as the authors suggest, that, "What impact these contacts have on attitudes . . . depends on the nature of the contact rather than its mere occurrence" (1951:71).

It would seem that the pretest–posttest paradigm was probably the strongest paradigm feasible in this situation, as there apparently were no other projects in the area which could have provided a comparison group. (Had such a comparison group been available, the study might have utilized the more powerful nonequivalent control group paradigm.) Since the pretest–posttest paradigm is employed,

TABLE 3.10 PERCENTAGES OF "HOSTILE" ATTITUDES ON THE THREE MEASUREMENT OCCASIONS

Attitude measurement occasion	Percentage expressing attitude that "Neighbors are low class"	Percentage expressing attitude that "Neighbors are not cooperative"
1st	21	34
2d	22	35
3d	32	38

* Adapted from Festinger and Kelley 1951:70. Used by permission of Research Center for Group Dynamics, Institute for Social Research, University of Michigan.

extraneous variables pose a substantial problem in this study, as in most studies where this paradigm is used. Any extraneous variable that changed over the same time period as the independent variable (increased contact) must be suspected of influencing the dependent variable (in this case, negative attitudes). The problem, of course, grows more serious as the length of time between the measurements increases. Since the time between measurements was approximately two and one-half months for both pretest-posttest studies involved in this research, a considerable number of extraneous variables could have changed during these intervals. It is easy to envision extraneous variables that might have influenced the outcome—for example, if there had been a substantial increase in burglary or other crime in the project during the critical periods, this might have increased the residents' negative attitudes.

The problem of extraneous variables may be lessened by careful monitoring of events that occur during the interval between measurements. If, as in this case, the researchers are able to continuously observe the research situation, it is likely that major changes in extraneous variables such as crime rate will be noticed. If the researcher has reason to expect that one or several particular extraneous variables may prove troublesome, he can control for this source of error by arranging to measure these variables. If he can show by these measurements that the suspect extraneous variables have not, in fact, changed during the period of the research, he can rule them out as rival hypotheses. If control through measurement of extraneous variables is not possible, the researcher should at least try to keep a careful log of occurrences during the interval between measurements, so as to maximize the chances that he will be aware of any changes in extraneous variables which might provide rival hypotheses to the one he is testing.

As noted in Chapter 2, the pretest-posttest paradigm also is subject to error from reactive measurement effects. You will recall that we have defined reactive measurement error as change in the dependent variable due solely to the measurement process itself. Although the questions in the structured interview used to assess hostile attitudes were carefully phrased in a neutral manner, there is at least a possibility that the interview itself might have acted subtly to change people's views about their fellow project residents. The researchers in this study, however, employed a method of sampling which reduced the possibility of error due to reactive measurements. Instead of interviewing exactly the same sample of project residents in the three surveys, they selected a new random sample of residents for each survey, with the result that the majority of those interviewed in the second and third surveys had not been interviewed before. This procedure should have considerably reduced any effects from reactive measurements.¹³

We can make the following observations regarding the reliability and validity of the measure of hostile attitudes, the dependent variable. The scoring of interview responses for "hostile expression" was done, we are told, in accordance with a content analysis scheme developed specifically for this study. No details of the coding scheme are given, however, and no information on interrater reliability is

¹³ It is possible to guard completely against reactive measurement effects in a pretest-posttest paradigm by taking measurements from two entirely separate random samples of informants—that is, by insuring that none of the persons included in the pretest random sample are included in the posttest random sample (cf. Campbell and Stanley 1966:53-54).

provided. Festinger and Kelley do comment that, "to insure a continuing high level of agreement between coders, periodic group discussions of specific problems in the code and constant supervision and check coding were instituted" (1951:80). It would appear from this that a satisfactory level of interrater reliability was achieved, but specific information on the degree of this reliability is needed. Similarly, no formal evidence for the validity of the content analysis measure of hostile attitudes is provided. It might reasonably be argued, however, that this method possesses considerable face validity. Certainly, responding to the rather bland interview questions used—for example, "What things about the project make it easy or hard to have community activities?"—with expressions of hostility toward one's fellow project residents would seem, on the face of it, to indicate a certain degree of hostility.

The replicability of the study would have been considerably enhanced if we had been supplied with details of the procedures used in the content analysis of the interviews, and of the various community activity programs. Lacking this information, it would be difficult to replicate the study. Comprehensive information was provided, however, on other important aspects of the study, including sampling techniques and the interview used.

THE NONEQUIVALENT-CONTROL-GROUP PARADIGM

Achievement Training in India

We have already discussed two studies dealing with achievement motivation. One, McClelland's 1961 study, suggested that the level of achievement motivation in a society is an important variable for the study of economic development, since the average level of achievement motivation in a society appears to partly determine its rate of economic growth. The second study, by LeVine, hypothesized that one of the determinants of the level of achievement motivation in a society is the nature of its "status mobility system." Recently, a great deal of attention has been paid to other possible determinants of achievement motivation. Researchers have, in particular, begun to explore the question of whether it might be possible for a person to increase deliberately his own personal level of achievement motivation. If a practical means of doing so is found, it might become possible for large numbers of persons in a society to voluntarily raise their level of achievement motivation, resulting, perhaps, in their society's (as well as, possibly, their own) rate of economic development being accelerated.

Accordingly, McClelland and Winter (1969) developed the hypothesis that achievement motivation might be increased by a specific course of instruction which included the following types of training inputs:

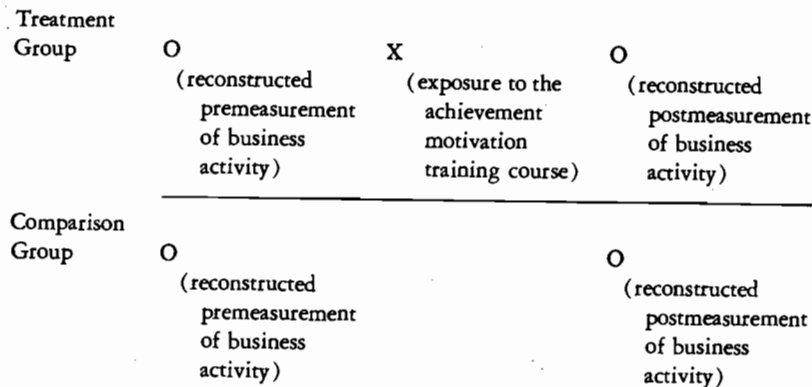
Teaching participants how to recognize and create achievement-oriented fantasy;
Encouraging them to consider how achievement relates to their own life-situations;
Guiding them in setting specific achievement goals and establishing guidelines by which their success may be judged;
Providing interpersonal supports for changes in motives.¹⁴

¹⁴ A full description of the components of the course may be found in McClelland and Winter (1969:45-78).

In order to test their hypothesis, McClelland and Winter designed a study that employed the nonequivalent-control-group paradigm. They arranged to make the course they had designed available at a research institute in India. They then contacted a number of businessmen from several Indian cities and after fully explaining the aims of the course, invited them to participate. A number of these men volunteered for the course and went ahead to complete it. These men constituted the treatment group. To obtain a comparison group, they contacted other businessmen who resembled the volunteers in important characteristics and were from the same cities.

A number of different measures of the dependent variable of achievement motivation were constructed. The basic measure used, the one which will be reported here, was a Business Activity Level scale developed for the study. Since all the participants were businessmen, it was expected that any increases in achievement motivation would be reflected in an increase in their business activity. The researchers therefore developed a set of criteria by which a man's business activity level could be scored. The scale ranged from -1 , indicating a low level of activity in business, to $+2$, indicating the man was highly active. Interviews were conducted with the members of the treatment and comparison groups about two years after the conclusion of the course and information was collected on the personal and business activities of each. On the basis of this information, the business activity level of each of the men was scored for the two years prior to the time of the course (the years 1962-1964) and the two years following (1964-1966). (It should be emphasized that these measures of business activity were *reconstructed* from later information provided by informants.)

The McClelland and Winter study may be diagrammed as follows:



The results of the study, in terms of business activity levels, are presented in Table 3.11. The treatment group, as predicted by the hypothesis, showed a marked increase in the percentage of persons who could be classified as active ($+2$) on the business activity scale. The comparison group showed only a very slight increase. A test of statistical significance indicated that it was very unlikely that these

TABLE 3.11 PERCENTAGE OF TREATMENT AND COMPARISON GROUP SUBJECTS CLASSIFIED AS ACTIVE (SCORING $+2$ ON THE BUSINESS ACTIVITY SCALE) DURING TWO YEAR PERIODS

	Before course 1962-1964 (percent)	After course 1964-1966 (percent)	Net percentage increase
Treatment Group	18	51	33
Comparison Group	22	25	3

Adapted from McClelland and Winter 1969:213, Table 7.2. Copyright © 1969 the authors. Used by permission.

differences were due to chance, since the probability of obtaining such great differences through chance alone was less than one in a thousand.

The use of the nonequivalent-control-group paradigm in this study means that this research is relatively invulnerable to one source of error that would plague the weaker static-group-comparison paradigm used in a similar situation. This is the rival hypothesis that the difference between the groups simply reflects pre-existing differences in the business activity levels of the comparison and treatment groups because of the manner in which the members of the group were selected. Thus, had a simple static group comparison been employed it might be maintained that the selection process—namely, using volunteers for the treatment group, while the comparison group was composed of nonvolunteers—had caused men with higher levels of business activity to be included in the treatment group. This would assume that more active businessmen will be more likely to volunteer for such a course. If, because of selection, the men in the treatment group were more active to begin with—that is, prior to their exposure to the independent variable—a static group comparison would show the two groups to be different even if the men simply maintained their original activity scores and the independent variable had no effect. A nonequivalent-control-group paradigm, such as the one used here, can distinguish between preexisting differences due to selection and differences which develop following exposure to the independent variable since it is the gain scores (the differences between the premeasurements and postmeasurements) that are compared and not the postmeasurement differences. Therefore, in this research we can reject the rival hypothesis of preexisting differences due to selection.

The only danger from selection when a nonequivalent-control-group paradigm is used is the possibility of interaction effects involving selection. Thus, it is conceivable that the volunteers who make up the treatment group in this study differed from the nonvolunteers in being more open to the exploitation of new business opportunities. If important new business opportunities developed during the study, the treatment group members might have taken greater advantage of them thereby increasing their level of business activity relative to the comparison group members. Such an interaction between selection and the extraneous factor of new business opportunities could cause the treatment group to show a significant gain relative to the comparison group, even if the independent variable had no effect. We can-

not reject this particular rival hypothesis in the present study, and it may be possible to develop other rival hypotheses involving selection interactions as well.

The possibility of interaction effects involving selection would have been greatly reduced, if not totally eliminated, if McClelland and Winter had found it possible to draw the members of the comparison group from among those who volunteered to take the course—that is, if they had randomly assigned volunteers to take the course or not take it. Had this been done, the study would have employed the control-group paradigm, which is stronger than the nonequivalent-control-group design precisely because it is resistant to selection interaction effects.

A major question concerning the McClelland and Winter study pertains to the validity of their measure of business activity. It will be recalled that all the data on a subject's business activity were collected from the subject himself roughly two years after he had completed the course. Measurements of this sort are subject to systematic distortion that may itself constitute a rival hypothesis to the one being tested. It might be, for example, that those who went through the course of training feel some need to demonstrate that it has been effective. Thus, as McClelland and Winter themselves note (1969:208–209), informants may tend consciously or unconsciously to distort their recollection of past business experiences in order to give the impression of improvement. Such distortion could create the appearance that the independent variable had had an effect when it in fact had none. Alternatively, some of the individuals in the comparison group who knew of the training program might have felt a need to demonstrate that they had made just as much progress as those in the treatment group and, as a result, might have consciously or unconsciously enhanced their recent business activity. Such distortion could obscure a real effect on the part of the independent variable. This source of error could have been reduced or eliminated by using measurements that were not so subject to memory error and intentional distortion. For example, as a substitute, the researchers might have interviewed the associates of the participants in the study both before the independent variable was introduced and afterwards, to ascertain the participants' levels of business activity. If this had been done there might have been less possibility of memory error and/or intentional distortion.

In sum, the validity of the measurements in this study is compromised by the dependence on reconstructed measures. Moreover, the interrater reliability for the rating system used to assign business activity scores is not reported, although an earlier version of the scoring system was said to have exhibited the relatively high interrater reliability of .92.

The replicability of the study is generally good, as the authors are careful to supply details of the various procedures employed. The study also incorporates provisions for dealing with the role of chance, since, as mentioned above, an appropriate statistical test was employed.

Innovation Adoption in Puerto Rico

An interesting anthropological study using the nonequivalent-control-group paradigm was carried out by Suchman, Ceollero, Munoz, and Pabon (1967). Their

study dealt with the acceptance of an innovation, a topic currently receiving considerable attention in applied anthropology. The innovation in question was a newly developed glove designed to protect cane-cutters' hands from injury due to misdirected machete blows, and from sores and skin irritation due to contact with cane stalks. In Puerto Rico, where the study was done, such injuries have constituted a major and continuing problem for cane-cutters.

One hypothesis to be tested by the study was that acceptance of an innovation, such as the protective glove, could be effectively promoted by bringing community pressure to bear on the intended recipients, in this case the cane-cutters. The idea was to enlist the support of community leaders in a campaign to promote the innovation. This, it was hypothesized, would induce the cane-cutters in the community to accept the glove. The rationale for this was that "since the individual shares the needs and attitudes of his sociocultural system, it is to be expected that if the group accepts an innovation, individual members will tend to accept it also" (Suchman et al. 1967:215).

This hypothesis may be paraphrased as follows:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
Support of community	will promote	adoption of an innovation.

To test this hypothesis, two "reasonably comparable" communities were selected for study. In one community, a strong effort was made to enlist the support of the community leaders for a campaign to promote the use of the protective glove. These leaders agreed to participate and several community meetings were held to discuss the campaign. A meeting was also held for the wives, mothers, and daughters of cane-cutters.

A second community provided a comparison group. Here no attempt was made to enlist the support of community leaders. Instead, the emphasis was placed on exposing the cane-cutters directly to educational material—including pamphlets and films—advocating use of the glove.

Subsequently, protective gloves were made available in both communities to workers who asked for them. Records were kept of how many workers accepted the glove and of how long those who accepted the gloves continued to use them.

This study, as was suggested above, can be viewed as employing the nonequivalent-control-group paradigm. This is true in spite of the fact that no premeasurements were obtained by Suchman et al. Our reason for classifying this study as using the nonequivalent-control-group paradigm is that it can legitimately be assumed, in this sort of situation, that prior measurements would have shown a zero level of the dependent variable (adoption of an innovation—in the present instance, a protective glove). Since, in this study, the innovation was unavailable to both groups until after the treatment group had been exposed to the independent variable, premeasurements, had they been made, would necessarily have shown no acceptance. The study may be diagrammed as follows:

Treatment Group (cane-cutters in the community where the community-pressure approach was used)	O (premeasurement of dependent variable of adoption of glove assumed to be zero)	X (enlisting support of community leaders for use of the glove)	O (whether or not worker accepted and continued to use glove)
Comparison Group (cane-cutters in the community where the direct educational approach was used)	O (premeasurement of dependent variable of adoption of glove assumed to be zero)		O (whether or not worker accepted and continued to use glove)

It was found that the rate of adoption of the glove (defined as accepting and continuing to use the glove for the duration of the study) was 19 percent in the treatment group, as compared to 76 percent in the comparison group. Obviously the community pressure hypothesis does not fare very well. The authors suggest that the following might have been among the factors involved in its failure. The community which provided the treatment group of cane-cutters was in a "state of flux" and the leadership structure was "diffuse and divided" (Suchman et al. 1967: 220). Under these conditions it is perhaps not surprising that community pressure was relatively ineffective in promoting acceptance of the innovation.

It should be stressed that if it could not safely be assumed that premeasurements would show a zero rate of adoption—which would have been the case, for example, if some protective gloves had been in use before the research began—this study would not represent an example of the nonequivalent-control-group paradigm. It would instead constitute an instance of the weaker static-group-comparison paradigm. Then the rival hypothesis of selection, which is ruled out by the nonequivalent-control-group paradigm, would have to be seriously considered.

The reliability and validity of the measure of the dependent variable in this study, adoption of the protective glove, would appear to be very high indeed, as the measure consists simply of observing whether or not the worker accepts and continues to use it. The replicability of the study is good as full details are provided of the procedures employed. No formal means were used to assess the role of chance, but it appears from examination of the results that a standard test of significance would show a very low probability of the results being due to chance.

CONTROL-GROUP PARADIGM

The control-group paradigm, since it requires random assignment of cases to treatment and comparison groups, is perhaps rarely feasible in the naturalistic research settings of interest to most anthropologists. However, studies by Lefkowitz,

Blake and Mouton (1955), Berelson and Freedman (1964), Freedman and Takeshita (1969), and Takeshita (1966) demonstrate how this powerful paradigm may sometimes be effectively applied under natural conditions.

Status, Models, and Violation of Prohibitions

Two hypotheses of interest to Lefkowitz and his coworkers in their 1955 study were: (a) people will be more likely to violate a prohibition when they have witnessed another person violating it; and (b) if the person who is seen to violate the prohibition (the "model") is of high status, he will be imitated more often than if he is of low status:

INDEPENDENT VARIABLE	NATURE OF RELATIONSHIP	DEPENDENT VARIABLE
a) Witnessing a violating model	increases	tendency to violate a prohibition.
b) Witnessing a violating model of high status (as opposed to one of low status)	will result in a greater increase in	tendency to violate a prohibition.

To test these hypotheses, the researchers focused on a natural setting, three street corners in Austin, Texas. This setting contained prohibitions in the form of traffic signal lights that alternately signaled pedestrians to "wait" or "walk." Two "models" were employed by the researchers. One was dressed in a manner which, in that area and at that time, implied high status—he wore "a freshly pressed suit, shined shoes, white shirt, tie, and straw hat." The second model was dressed in a manner to imply low status—"well-worn scuffed shoes, soiled patched trousers and an unpressed blue denim shirt." Actually, both models were portrayed by the same person, a man in his early thirties who made the transition from one model to another simply by changing his clothes.

The procedure used was for either the high- or the low-status model to cross the intersection during the middle of the "wait" signal. Pedestrians who happened to be standing with the model before he crossed the street were then observed from a distance to see whether they imitated the model in disobeying the prohibition. A number of trials were made at the various street corners over a period of three days. Observations were also made during a comparable number of intervals during which no model was present in order to ascertain the frequency of violations in the absence of a model. The timing of the various trials was carefully arranged to avoid any consistent relationship between time of day and type of trial.

This study can be said to employ the postmeasurement-only version of the control-group paradigm (see p. 17, footnote 10), since it can reasonably be maintained that the appearance of individual pedestrians at a street corner during a particular interval

of time is essentially a random process. Thus, the researchers took advantage of a kind of natural randomization process to achieve something close to random assignment of cases to treatment and comparison groups. This made it possible to assume that the two groups were equivalent prior to exposure of the treatment group to the independent variable.

The test of the first hypothesis, that people will be more likely to violate a prohibition when they have witnessed a model violating it, can be diagrammed in our usual manner, as follows:

Treatment Group (pedestrians witnessing a violating model)	R (natural random assignment of pedestrians)	X (witnessing models—both high and low status—violate the prohibition on crossing the street)	O (whether pedestrians violated the prohibition)
Comparison Group (pedestrians not witnessing a violating model)	R (natural random assignment of pedestrians)		O (whether pedestrians violated the prohibition)

The results for this first hypothesis are summarized in Table 3.12. The prediction that violations will increase with exposure to a violating model is strongly supported, there being an increase from 1 percent violations for the comparison group

TABLE 3.12 THE EFFECT OF THE PRESENCE OF A MODEL ON PEDESTRIAN VIOLATIONS

	<i>Pedestrians obeying prohibition</i>	<i>Pedestrians violating prohibition</i>	<i>Percentage of violators</i>
Exposed to Violating Model (includes both high and low status models)	526	52	9
Not Exposed to Violating Model	742	8	1

Adapted from Lefkowitz et al. 1955:705, Table 1. Copyright 1955 by the American Psychological Assn. Used by permission.

to 9 percent for the treatment group. A statistical test showed that this difference was very unlikely to have been due to chance. The test of the second hypothesis can be diagrammed as follows:

Treatment Group (pedestrians witnessing a high status model)	R (natural random assignment of pedestrians)	X (exposure to a high rather than low status violating model)	O (whether pedestrians violate the prohibition)
Comparison Group (pedestrians witnessing a low status model)	R (natural random assignment of pedestrians)		O (whether pedestrians violate the prohibition)

Table 3.13 presents the data relevant to this hypothesis. The second prediction from this hypothesis (that exposure to a high-status model will result in a greater tendency to violate a prohibition than will exposure to a low-status model) is also sustained. More than three times as many pedestrians violated the signal after observing a high-status violating model than did so after witnessing a low-status violating model. And again, as in the case of the first hypothesis, a standard statistical test showed that this difference was very unlikely to have been due to chance.

TABLE 3.13 THE EFFECT OF ROLE MODEL STATUS ON PEDESTRIAN VIOLATIONS

	<i>Pedestrians obeying prohibition</i>	<i>Pedestrians violating prohibition</i>	<i>Percentage of pedestrians violating prohibition</i>
Exposure to High-Status Violating Model	250	40	14
Exposure to Low-Status Violating Model	276	12	4

Adapted from Lefkowitz et al. 1955:705, Table 1. Copyright 1955 by the American Psychological Assn. Used by permission.

There is relatively little to criticize concerning this study. The paradigm used, the control-group paradigm is a very strong one. There is, perhaps, one point of uncertainty concerning the randomization process. It is only an assumption that the arrival of pedestrians at an intersection during a particular interval is essentially a random process. However, there seems to us little reason to question such an assumption.

The reliability of the measurements used seems likely to be high since the measurement process consisted simply of an observer's recording violations or non-violations by pedestrians. The measure also appears to be valid since it is actually a sample of the behavior which it is desired to measure—violation of a prohibition. The study is described in sufficient detail to provide for easy replication, and statistical tests were employed to ascertain that the differences obtained were unlikely to be due to chance alone.

Fertility Control in Taiwan

Population growth is a serious issue of great practical importance in all parts of the world today. Like the weather, however, it appears that more people are talking about it than are doing something effective about it. To be sure, many groups, agencies, foundations, and countries are working to alter the potentially disastrous trend toward rapid population growth. However, the question is, "How effective are these efforts?" Many programs can point to changes in birth rates or contraceptive-use rates, but it is not always clear that these differences are due to the program making the claim.

A recent study of a program to increase the percentage of Taiwanese women using contraceptive devices (and in this way to lower the birth rate) provides a useful illustration of how the control-group paradigm may be employed to test hypotheses concerning the efficacy of such programs of directed change. This study thus has particular relevance for applied anthropology. The following summary of this research is based on reports by Takeshita (1966), Berelson and Freedman (1964), and Freedman and Takeshita (1969).

Preparatory research was begun in 1962 before the start of the family-planning campaign. At the time, Taiwan's population was 12,256,682, distributed over 14,047 square miles of land (an average of 873 persons per square mile). Birth rates were declining slowly but death rates were declining rapidly. The result was a steady annual growth in population of 3 percent. The initial research indicated the vast majority (about 90 percent) of women in the prime reproductive ages of 20-39 wanted to limit the size of their families but did not know how to do so effectively. Generally, they were anxious to learn about more effective means for doing so, especially if they already had a few children, including at least one son. The goal of the campaign then became one of transforming existing attitudes into action rather than of convincing women that they should desire to have fewer children.

The campaign itself was begun after this prestudy had been completed, and the large city of Taichung was chosen as the site. The program was massive and complex. Berelson and Freedman suggest that it also constituted "one of the most

extensive and elaborate social science experiments ever carried out in a natural setting" (1964:32).

In Taichung there were 36,388 married women in the prime reproductive age group out of a population (in 1963) of about 300,000 people. Four basic approaches were utilized to communicate information about birth control devices to these women. The researchers designated them as "Everything: wives and husbands," "Everything: wives only," "Mail," and "Nothing." The "Everything: wives and husbands" approach included posters, mailed literature, mass meetings, small group meetings,¹⁵ and visits by public health nurses to both husbands and wives. The "Everything: wives only" approach was the same as the preceding except nurses visited only the wives. The "Mail" approach consisted of mass meetings, mailing of family-planning literature, and the use of posters. The "Nothing" approach involved only mass meetings and posters.

The city of Taichung is made up of 2,389 neighborhoods or *lin's*, each having from twenty to thirty families. These *lin's* were the basic units in the research. Approximately one third of the *lin's* received the "Nothing" treatment, another one third received the "Mail" treatment, approximately one sixth received the "Everything: wives only" treatment, and the remaining one sixth received the "Everything: wives and husbands" approach. The *lin's* were assigned to these various treatment conditions at random, although in the proper proportions.¹⁶

A basic hypothesis implicit in the study was that the more intense the campaign, the greater would be the rate of acceptance of family planning assistance. This may be broken down into six subhypotheses, as follows:

1. The "Everything: wives and husbands" approach would increase acceptance relative to the "Everything: wives only" approach;
2. The "Everything: wives and husbands" approach would increase acceptance relative to the "Mail" approach;
3. The "Everything: wives and husbands" approach would increase acceptance relative to the "Nothing" approach;
4. The "Everything: wives only" approach would increase acceptance relative to the "Mail" approach;
5. The "Everything: wives only" approach would increase acceptance relative to the "Nothing" approach;
6. The "Mail" approach would increase acceptance relative to the "Nothing" approach.

The basic paradigm employed in the study was the posttest-only version of the control-group paradigm. The tests of the six subhypotheses may be diagrammed as shown on pages 76-78.

¹⁵ Small group meetings were inadvertently omitted in some cases.

¹⁶ The random assignment of *lin's* to treatments was subject to certain geographical constraints (see Berelson and Freedman 1964).

Sub-hypothesis 1

Treatment Group

(<i>lin's</i> which received the "Everything: wives and husbands" treatment)	R (random assignment of <i>lin's</i> to Treatment Group)	X (receiving the "Everything: wives and husbands" treatment rather than the "Everything: wives only" treatment)	O (percentage of acceptances)
---	---	--	----------------------------------

Comparison Group

(<i>lin's</i> which received the "Everything: wives only" treatment)	R (random assignment of <i>lin's</i> to Comparison Group)	O (percentage of acceptances)
---	--	----------------------------------

Sub-hypothesis 2

Treatment Group

(<i>lin's</i> which received the "Everything: wives and husbands" treatment)	R	X (receiving the "Everything: wives and husbands" treatment rather than the "Mail" treatment)	O
---	---	--	---

Comparison Group

(<i>lin's</i> which received the "Mail" treatment)	R	O
---	---	---

Sub-hypothesis 3

Treatment Group

(<i>lin's</i> which received the "Everything: wives and husbands" treatment)	R	X (receiving the "Everything: wives and husbands" treatment rather than the "Nothing" treatment)	O
---	---	---	---

Comparison Group

(<i>lin's</i> which received the "Nothing" treatment)	R	O
--	---	---

Sub-hypothesis 4

Treatment Group

(<i>lin's</i> which received the "Everything: wives only" treatment)	R	X (receiving the "Everything: wives only" treatment rather than the "Mail" treatment)	O
---	---	--	---

Comparison Group

(<i>lin's</i> which received the "Mail" treatment)	R	O
---	---	---

Sub-hypothesis 5

Treatment Group (<i>lin's</i> which received the "Everything: wives only" treatment)	R	X (receiving the "Everything: wives only" treatment rather than the "Nothing" treatment)	O
--	---	---	---

Comparison Group (<i>lin's</i> which received the "Nothing" treatment)	R		O
--	---	--	---

Sub-hypothesis 6

Treatment Group (<i>lin's</i> which received the "Mail" treatment)	R	X (receiving the "Mail" treatment rather than the "Nothing" treatment)	O
--	---	---	---

Comparison Group (<i>lin's</i> which received the "Nothing" treatment)	R		O
--	---	--	---

The data relevant to these subhypotheses are summarized in Table 3.14. It can be seen that increased effort did not necessarily enhance the effectiveness of the program. Thus, the use of letters did not increase the acceptance rate for family-planning services and contacting both husband and wife was no more effective than contacting only the wife. However, one specific type of approach emerges as highly effective. Home visits by health workers to wives together with neighborhood small group meetings (the "Everything: wives only" treatment) did produce a marked increase in the utilization of family-planning services. Findings such as these are obviously of great practical importance as they enable those in charge of family-planning programs to channel limited funds into those activities most effective in promoting program goals.

TABLE 3.14 ACCEPTANCE RATES FOR CONTRACEPTIVE METHODS PER 100 MARRIED WOMEN AGED 20-39 IN TAICHUNG UP TO APRIL 1, 1964^a

Treatment	Number of acceptors per 100 women
Everything: husband and wife	17
Everything: wife only	17
Mail	8
Nothing	8

^a Adapted from Freedman and Takeshita (1969:126). Table VI-3. Reprinted by permission of Princeton University Press.

This research, of which only a small part has been summarized here, is impressive, both in its scope and in the general quality of its design and execution. The use of the control group paradigm makes it possible to discount a large number of rival hypotheses since it provides effective control over the major error sources.

There would seem to be no serious reliability and validity problems in the study. The dependent variable, acceptance of family-planning assistance, is operationally defined simply as "the insertion of an intrauterine device or the receipt of instructions and the purchase of supplies for other methods, together with expressed intent to practice contraception" (Berelson and Freedman 1964:34). Due to its simplicity and unequivocal nature, the measure seems likely to possess satisfactory reliability. It also features a very high degree of face validity.

The replicability of the study seems generally high, although it would be still further enhanced if additional details of the techniques used were made available—for example, the contents of the mailed materials and the format of the meetings.

A limitation of the data analysis in this study is the failure to make explicit use of statistical analysis in order to deal with the role of chance. While the observed differences between groups are of such a magnitude that it appears extremely unlikely they are due to chance, it would seem desirable to apply statistical tests of significance.

SUMMARY

As we have indicated previously in this chapter, hypothesis-testing research in anthropology has overwhelmingly relied, to date, on the static-group-comparison paradigm. More powerful paradigms that require premeasurements, such as the pretest-posttest and nonequivalent-control-group paradigms, have been quite rare.¹⁷

The prominence of the static group comparison in anthropological research appears to be the result of several factors. One reason is that many of the processes of interest to anthropologists are relatively long-term—that is, they operate over time spans of several human generations or longer. Examples of such long-term processes might include changes in kinship systems and religious ideology, adapta-

¹⁷ One powerful paradigm, the postmeasurement-only version of the control-group paradigm, does not, of course, require premeasurements. It and the standard control-group paradigm do, however, require random assignment of cases and thus are unlikely to become widespread in anthropology, at least for the time being.

tion of cultural institutions to ecosystem features, or changes in child-training practices. When one attempts to test hypotheses involving very long term processes, there is perhaps little alternative to the use of a static group comparison since the premeasurements required in other paradigms often are not feasible. Another and related reason for the predominance of the static group comparison in anthropology is that the nature of certain sociocultural processes precludes premeasurements. We saw an example of such a process in Wolf's research on the relationship between infantile association and sexual attraction. A premeasurement paradigm would seem to have been unfeasible here due to the impossibility of obtaining a premeasure of sexual attraction from infants. When dealing with long-term processes and processes not amenable to premeasurement, then, the more powerful paradigms are generally precluded and the use of the static group comparison is unavoidable. As we have pointed out in reviewing the research by LeVine (1966) and Wolf (1970), special procedures may be used in many cases to partially compensate for the weaknesses of the static group comparison.

However, when we exclude from consideration long-term processes and processes inherently not amenable to premeasurement, we are left with a large number of intermediate and short-term sociocultural phenomena that would seem to lend themselves to study using powerful paradigms incorporating premeasurements. It is surprising, for example, that such paradigms have so rarely been used in studies of the short-term aspects of acculturation, modernization, and technological change, to give only a few of the more obvious examples.¹⁸

One of the more important reasons for the underutilization of the more powerful paradigms would seem to derive from existing conventions of anthropological field research. Field research in anthropology is conventionally done by means of relatively short-term, "one-shot" projects, often with a duration of a year or less. If premeasurement paradigms are to be applied, the research must encompass a long enough time span to "capture" the process of interest, that is, to permit measurements to be made both "before" and "after." Very many of the "short-term" processes of interest to anthropologists still have time spans measured in years rather than months or days—for example, the process of personality change that occurs in individuals who have moved from villages to cities. The study of phenomena such as this will obviously call for a lengthy research program that might take the form of an initial "baseline" study to obtain premeasurements and one or more additional studies over a period of several years in order to obtain postmeasurements. As long as anthropology continues to be wedded to the idea of the short-term, "one-shot" research program, the opportunities to apply powerful premeasurement paradigms will be severely and detrimentally circumscribed.

Of the paradigms requiring premeasurements, the pretest-posttest paradigm offers, perhaps, the smallest margin of benefit over the static group comparison.

¹⁸ The reader familiar with the anthropological literature will be aware that the pretest-posttest paradigm has been employed to some extent in anthropological research on these topics. Generally, however, such studies have substituted for an actual premeasurement a *reconstructed premeasurement* based on informants' recollections of past events, historical documents, or, in some extreme cases, simply the researcher's imaginative powers. With certain exceptions, such "reconstructed" premeasures are extremely error-prone and are very difficult to quantify. For these reasons this version of the paradigm is not recommended.

This paradigm, while it avoids error due to selection, is quite subject to the effects of extraneous variables and reactive measurement effects. In the pretest-posttest paradigm, rival hypotheses concerning extraneous variables may be ruled out if the extraneous variables can be measured (as was shown above in our review of Festinger and Kelley's research). If, however, a great many extraneous variables are changing during the same period as the independent variable, it will be very difficult, if not impossible, to rule out rival hypotheses involving them. In such cases, the advantages to be gained from using this paradigm in preference to a simple static group comparison may be minimal.

Much to be preferred over either the pretest-posttest paradigm or the static group comparison is the nonequivalent-control-group paradigm. It not only is resistant to simple selection errors and reactive measurement effects, but also provides a great deal of control over extraneous variables. In situations where a pretest-posttest paradigm is feasible it is very often possible, with relatively little additional effort, to obtain a suitable comparison group as well, thus permitting the use of the nonequivalent-control-group paradigm. The advantages of this latter paradigm are so pronounced that it becomes virtually mandatory that the pretest-posttest paradigm be used only in situations where, for some reason, the nonequivalent-control-group paradigm is infeasible.

It is perhaps unlikely that the strongest of our paradigms, the control-group paradigm, will play a major role in anthropological research, at least in the near future; it is probable that anthropological research will continue to be characterized by commitment to the study of natural, on-going social systems where investigator intervention (which this paradigm generally requires) is undesirable and impracticable. The Lefkowitz et al. study (1955) does demonstrate, however, that it is sometimes possible to carry out research using a control-group paradigm within a natural social setting. And the study of fertility and birth-rates in Taiwan by Berelson and Freedman demonstrates a use of this paradigm that involved only minimal and relatively routine intervention by the investigators. Thus, it behooves anthropologists to be on the alert for situations in which this powerful paradigm may be applied in their research.