Theoretical Statements Relating Two Variables

We said in the previous chapter that a hypothesis must be made up of at least two variables; a hypothesis expresses the relationship between two or more variables. We also referred to assumptions and propositions. In this chapter, we look at the relations between two variables more closely and distinguish among several types of theoretical statements.

Hage (1972) suggested that the term theoretical statement be used to describe more broadly the relationship between variables, encompassing such terms as assumption, hypothesis, postulate, proposition, theorem, and axiom. In this book, we will use three of the more specific terms—assumption, hypothesis, and proposition—with theoretical statement being used generally to describe all three.

A theoretical statement says something about the values of one or more variables, although it is generally thought of as expressing something about the relationship between two or more variables (e.g., “The more television a child sees, the more aggressive the child will act”). Other kinds of theoretical statements describe the values of only one variable (e.g., “There is a consistently high amount of violence on prime-time television”).
We find it useful to distinguish among three types of theoretical statements:

*Hypothesis*—A testable statement about the relationship between two or more concepts (variables); for example, “The more politically active people are, the more time they spend reading a daily newspaper.” *Testable* means that social science research methodology and statistics can be applied to discover the extent of support for the statement.

*Assumption*—A theoretical statement that is taken for granted, not tested. The assumption may describe the relationship between variables (similar to a hypothesis, above), or it may describe the usual value of one variable in a given situation (as in propositions, below). Some assumptions may be considered untestable or may be beyond the scope of the study: for example, “The more rational the electorate, the more it is motivated to seek political information in the mass media.” Others could be tested, but are taken for granted in a given study: for example, “The more politically active people are, the more motivated they are to get information about the election from the mass media.” Such assumptions are often used as theoretical linkages for hypotheses: that is, reasons why the hypothesis may be supported.

*Propositions* are less useful than theoretical statements that address the relationship between two or more variables because they are merely descriptive and provide information about only one variable at a time: for example: “The free flow of information is valuable in a democracy.” Propositions often take on a *normative* tone, in which scholars state how things should be, according to their ideological views.

**IDENTIFYING ASSUMPTIONS**

Whether assumptions are propositional or relational, it is crucial for the social scientist to specify as many assumptions underlying the research and theory as possible. Assumptions are necessary (not everything can be tested) and/or convenient (pragmatism requires that not every scholar go back to the ultimate cause, e.g., the big bang). All studies are based on one or more underlying assumptions. These form the logical rationale for the study and can be used to derive the hypotheses. Although the reader does not require, for example, an assumption of evolutionary biology when reading a study of most human behavior, it is helpful for scholars to clarify their own deeply held beliefs and to acknowledge these when directly pertinent to the study. The more scholars can identify the assumptions that underlie their theories and research, the more they and others can understand the implications of the theories. If the reader does not agree with the basic assumptions underlying the research, then the rest of the work is called into question. Therefore, identifying and communicating assumptions is a form of intellectual honesty. Sometimes assumptions are not or cannot be made explicit by the researcher, and it is up to the reader to identify them.

Unfortunately, identifying and stating assumptions is one of the most difficult parts of theory building because the things we take for granted are part of our individual ideological and normative systems and are therefore transparent to our daily thought processes. Such preconceived ideas have a direct effect on the research conducted. Consider ideas such as:

- Capitalism is the best economic system.
- Men are more capable than women are.
- The unemployed lack the intelligence or motivation to find and hold jobs.
- The United States operates as a democratic political system.
- The more information people can have about political candidates the better.
- Mass media content basically mirrors reality.
- Newspapers provide more information about an election than television news programs.

Many people may agree with such statements, but establishing informal agreement is not science. Most people used to agree that the sun orbited the earth, but agreement did not make it true, and it took scientific observation to show the opposite.

Science advances by testing of hypotheses, not by assuming that certain things are true. Assumptions can be challenged on logical or philosophical grounds, but in this case one person’s opinion may be as valid as another’s. Social science methods and statistics provide a potentially more objective form of evaluating the relative support for a hypothesis. These methods attempt to be independent of a given
scholar's personal biases and may be replicated by others. If they are replicated, we have more confidence in the original hypothesis test. Scientists recognize that no one is completely without bias and that too much bias can negatively influence the outcome of studies. Researcher bias is a threat to establishing internal validity: that is, showing that changes in the dependent variable are due to the independent variable and not to other causes.

As we indicated in the previous chapter, Hage (1972) suggested that the use of categorical variables may signal underlying assumptions that need to be specified. Why compare men's voting or newspaper reading to women's? What distinguishes men from women that makes us think that they will vote or read differently? Could there be assumptions about political interest, intelligence, and ability to understand abstract concepts, availability of time and transportation? If we can ask "why" when categorical variables are proposed, we may be able to uncover underlying assumptions, turn them into hypotheses, and actually test the extent to which assumptions are supported.

The lower the proportion of assumptions to hypotheses a theory has, the more it explains about the phenomenon at hand. Tested and supported hypotheses provide information about what the world is like. This is not the same as reaching truth, as Popper (1968, 1972) cautioned, but repeated hypothesis tests with consistent results do give us some reassurance about their validity. Assumptions, however, are mere guesses about the nature of the world and have no scientific support. What "everybody knows" may in fact be incorrect. Therefore, the fewer assumptions a theory contains relative to the number of hypotheses, the more power the theory has to describe, explain, and predict the world in a way than can be empirically defended.

FORMS OF HYPOTHESES

Hypotheses that express the relationship between two variables can be written in a wide variety of ways, some more intelligible than others. Basically, hypotheses tell us something either about the difference between the average values of variables (e.g., A is on the average bigger than B) or about the relationship between two variables' values (e.g., as the values of A change, the values of B also change).

A hypothesis of difference could take the following form: "Canadian exports to the United States account for a greater share of U.S. imports than the reverse." The independent variable is country, which has the values "Canada" and "the United States." The dependent variable is the proportion of imports, or, more specifically, the proportion of imports of each country that comes from the other. Such a hypothesis is testable using statistics that compare means, such as the t test and F test (analysis of variance).

Hypotheses that test relationships may take the form "if, then" or "the more, the more." For example, "If a country depends on another country for a large share of its imports, then its news media will include a lot of information about that country." This is not as useful a hypothesis form as it might seem, as we can see when we rephrase it into the more continuous form: "The more a country's imports come from another country, the more information its news media will include about the other country."

In the first form, the independent variable has two values, a "large" share and, we infer, a lesser share of imports. Likewise, the dependent variable has been dichotomized into a "lot" of information and less information. In contrast, the continuous relational form of the second hypothesis allows us to use both independent and dependent variables as continua, thus permitting the introduction into the study of countries with imports and information that vary in increments of all sizes.

It should be noted that the form of the relational hypotheses communicates direction of the relationship. "If A increases, then B increases" is a positive relationship because the values of the variables both change in the same direction. Likewise, "If A decreases, then B decreases" is also positive. The negative relationship, in which the variables' values change in the opposite direction, is expressed by "If A increases, then B decreases," or vice versa. The same is true of the continuous relational form: Positive relationships are expressed as "The more A, the more B" or "The less A, the less B." Negative relationships include "The more A, the less B" or "The less A, the more B." Most of the time, it is assumed that the first variable in the hypothesis, here A, is the independent variable, or cause, and the second variable is the dependent variable, or effect.

In general, we find the continuous relational form of hypotheses more useful than the categorical "If A, then B" form, for the same reasons that we preferred continuous concepts to categorical ones (see Chapter 2). As we see above, the categorical form "If A, then B" appears to provide less information than the continuous form
Figure 3.1 Both independent and dependent variables are continuous.

Hypothesis: The more education a person has, the more he or she reads a daily newspaper.
Operational definition for education: The number of years of formal schooling a person has.
Operational definition for newspaper readership: The number of days per week that a person inputs and processes information from a newspaper.

```
<table>
<thead>
<tr>
<th>Days per week read a daily newspaper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Years of education</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
</tbody>
</table>
```

"The more A, the more B." In fact, the operationalization of the concepts in the hypothesis defines whether the hypothesis is continuous or categorical.

Simply put, there are three combinations of categorical and continuous variables that yield three types of hypotheses:

1. Both variables are continuous. "The more people read a newspaper, the more interested they are in politics." The converse could also be supported because causal direction is ambiguous in this relationship. The hypothesis could also be phrased: "There is a positive relationship between political interest and newspaper reading." Figure 3.1 shows the relationship between individuals’ education and how frequently they read a newspaper. Both variables are continuous, allowing us to show what a hypothesized straight-line relationship would look like. The slope of the line is arbitrary, but, as we will learn in Chapter 4, the slope should approximate what the theory predicts.

2. One variable is categorical and one continuous. "Women vote more often than men do." Biological sex is assumed to be the independent variable, with number of elections voted in being dependent. This is a classic t-test statistic situation, with sex as the grouping variable and voting as the variable on which means are calculated for each value of sex. Figure 3.2 shows the same sort of relationship if we dichotomize education. The height of the columns is arbitrary but should approximate what the hypothesis predicts.

Although we often think of the grouping variable in a t test as being the independent variable, there are also hypotheses involving categorical and continuous variables in which causal direction is either ambiguous or not intended: for example, "Clinton supporters are younger than Bush supporters are." This hypothesis could certainly be tested with a t test, with candidate as the dichotomous grouping variable and age as the variable on which means are calculated. However, there is no implication that changing the candidate whom a person supports will
change that person’s age. Rather, the hypothesis is merely testing the difference in ages between the two groups. As always, making assumptions about causal direction is difficult and fraught with danger.

3. Both variables are categorical: “People with less education read the newspaper less [than people with more education].” As is often the case when dealing with categorical variables, not all values of the variables are always included in the hypothesis. Implicit is a continuation of the hypothesis: “than people with more education.” Two categorical variables may easily be analyzed with a contingency (cross-tabulation) table, with the presumed cause (in this case, political party) being the column variable and the presumed effect the row variable. Making the presumed cause or independent variable the column variable is a scientific convention only, but it is one that readers have come to expect. The interpretation of a table will be based on an assumption that this format is being used. Figure 3.3 further dichotomizes newspaper reading into high and low categories, producing a $2 \times 2$ contingency table. The number of Xs is intended to be an example of the number of people who would fall in each cell, according to the hypothesis.

In the next chapter, we will discuss operational linkages and learn more about how these hypotheses may be represented visually. One form of the operational definition is the graph, as in Figures 3.1 to 3.3.

❄ CAUSAL DIRECTION

In the first type of hypothesis defined above, to say, “If A happens, then B will happen,” implies that A is the cause and B the effect. The same is true of “The more A, the more B.” The assumption is that the first variable named is independent.

In cases when it is impossible or unwise to infer causal direction, scholars may wish to substitute the form “A is [positively or negatively] related to B.” This states the fact of the relationship and the direction of the relationship but does not imply causal direction. However, scholars should not use this more ambiguous relational form merely to avoid making statements that support their convictions. There are ways to argue causal direction that should be used between variables that are related. We caution scholars against using this form merely to avoid the task of establishing causal direction.

Figure 3.3 Both variables are categorical.

Hypothesis: People with less education tend to read the newspaper less [than people with more education]. (The part in brackets is implied, whether or not explicitly stated.)

Operational definition for education: The distribution in Table 3.1 (number of years of formal schooling) is dichotomized into high and low categories by splitting the distribution at the median. (Note: This is done for purposes of illustration only. In general, if you had a ratio-level operational definition, you would not want to dichotomize it. You would lose a substantial amount of information.)

Operational definition for newspaper readership: Take the distribution shown in Table 3.1 (days per week read a daily newspaper) and dichotomize it at the median. (Note: This is done for purposes of illustration only. In general, if you had a ratio-level operational definition, you would not want to dichotomize it. You would lose a substantial amount of information.)

<table>
<thead>
<tr>
<th>Education</th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Xxxxxx</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Xxxxxx</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Xxxxxx</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Xxxx</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Xxxx</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>Xxxx</td>
<td></td>
</tr>
</tbody>
</table>

Causal direction may be supported in four ways:

1. The hypothesis must show statistical support for covariation (i.e., a relationship) between the two variables: As one variable changes its values, the other’s values also change. Change can be in either a positive or a negative direction. In a positive relationship, the values of one variable change in the same direction as the values of the other. In a negative relationship, the values of the variables change in opposite directions.
2. The presumed cause should occur in time before the presumed effect. This is easy to establish in experiments, where the researcher has control over the timing of administering treatment variables to the subjects. In other research, such as cross-sectional surveys, however, it may be impossible to empirically establish time order, making causality more difficult to establish.

3. The researcher should rule out possible alternative explanations for the observed relationship. In an experiment, if subjects are randomly assigned to treatment groups, random assignment theoretically rules out variance due to an infinite number of unidentified and unmeasured variables. In surveys, however, researchers must use the literature to identify plausible variables that may be alternative explanations and then find ways to measure them and statistically control for them. The survey researcher’s task is more difficult but not impossible. The same constraints apply to content analysis.

4. All researchers must minimize error variance. This may be as simple as ensuring that no errors are made in the study—that all subjects get the designated stimulus, that all respondents are given the same question in the same tone of voice. In practice, however, errors creep in and are inevitable in research. The researcher’s job is to minimize them. Another form of minimizing error variance is controlling for important variables that are related to the dependent variable. In analysis-of-variance terms, this means keeping the error term of the equation as small as possible. The error term may include variability in the dependent variable not measured by the independent and control variables used in the study. The larger the error term, the less likely statistical significance is to be achieved.

How Research Questions and Hypotheses Differ

The appropriate use of hypotheses and research questions lies primarily in whether one is engaged in deductive or inductive research. The deductive model of science begins with theory, forms hypotheses, collects data to test the hypotheses, and then if necessary revises the theory. Inductive research begins with the data. It forms generalizations that become theory and may be later tested deductively. Although deductive research is ideal to test theories, inductive research is better at building theory.

Research questions are most appropriate in new areas of research in which little is known about the relationships among variables and in which there is scant literature that is applicable. Otherwise, hypotheses should be stated. With the explosion of social science research in recent decades, it is difficult to imagine a study proposal unrelated to any line of previous research. Although the topic may be new, as in the case of a new technology, there have been numerous studies about how people and society relate to new technologies, and surely these can suggest hypotheses for testing. Researchers should not avoid framing hypotheses merely because they are not sure whether they will be supported. Sometimes it is even more important to know that a hypothesis is not supported than that it is.

Research questions should be used only when there is a legitimate need for inductive theorizing. They should never be a substitute for a wide-ranging literature search and critical thinking on the part of the scholar. For example, “Does the nature of the protagonist affect how aggressive children are when they see televised violence?” is one or more hypotheses in disguise. The scholar should tap into the literature dealing with identification, fantasy versus reality of presentation, and so on, to form one or more hypotheses about this topic.

The primary danger with research questions is that the “answers” to the questions are often interpreted in the same way as hypothesis results. Yet hypotheses are interpreted narrowly. For example, in response to the question “Is there a difference between the amount that women and men read a newspaper?” it is certainly possible to observe that one mean is bigger than another. Let’s say we observe that women read newspapers more frequently. The researcher confidently reports the findings and confirms that women are more frequent readers than men are. Unfortunately, the only thing established is that in the sample more men than women read. But this is uninteresting information, for the purpose of most research (where less than the population is being studied) is to say something about the population, not about those individuals who by chance were included in the sample.

The same would be true if the means showed that men read newspapers more frequently. All we have established is that in this sample the means are as reported. But some researchers using research questions as confidently report one outcome as the other.

By contrast, the testing of a hypothesis requires that a direction be predicted. We look at the literature and find that in most studies men read newspapers more than women do. Thus, we frame the
hypothesis “Men read newspapers more frequently than women do.” Instead of using “eyeball statistics” to answer the research question, we conduct a $t$ test. The $t$ test gives us the advantage of knowing whether the observed difference between the means is large enough to represent (at some specified probability level) a real difference between the groups in the population or whether the difference is merely due to chance or random error.

Let’s say the $t$ test does support the hypothesis and, as in the research question, we conclude that men read newspapers more than women do. Statistical significance in the $t$ test implies that in the population men read newspapers more than women do, not that this is true just of the sample. This is a major advantage over the eyeball statistics used in the study with the research question.

But what if the results of the hypothesis test are different? What if the hypothesis is not supported? What if there is no difference between men’s and women’s reading, if women read a lot more, or if men read only slightly more? All three of these would result in lack of support for the hypothesis that men read more than women do. Does this mean that we can conclude that women read more than men do? Definitely not. The logic behind hypothesis tests requires that only statistically significant results in the direction of the hypothesis may be taken as supporting the hypothesis. All other results are ambiguous. Did we make a mistake in the study? Was the literature wrong? Further research may be necessary. Null results from a hypothesis test do not necessarily mean that the underlying theory is incorrect; there are many ways in which error may creep into studies, and to give the theory a fair test, the researcher must reconduct the study using different methods.

One last word about research questions: The scholar who poses research questions and then uses inferential statistics (such as the $t$ test) to answer them is committing an error of logic. Inferential statistics are for testing hypotheses, and the researcher should reformulate the research questions as hypotheses.

\section*{Notes}

1. Statistics books distinguish between null and research hypotheses. The null hypothesis is a statement of no difference between the values of a variable or no relationship between variables. Technically, the null hypothesis is tested by statistics. The research hypothesis states the opposite, but it is generally reported in research accounts. Statistical significance implies that the null hypothesis may be rejected and that there is a certain probability that the predictions of the research hypothesis can be generalized from the sample to the population. In this book, the term hypothesis can be assumed to mean the research hypothesis.

2. We use the term eyeball statistics to indicate instances where a researcher looks at, for example, the difference between two means and says, “Well, it looks like Mean A is bigger than Mean B.” But how big is big? How big a difference is enough to say something meaningful? Inferential statistics are always better than eyeball statistics, even if the two yield the same result. Inferential statistics allow us to estimate the probability of our being wrong when we say the two means are different. Eyeball statistics rely on a wing and a prayer.
Once hypotheses are formed, it is necessary to specify two sorts of linkages, connections among the variables in the hypothesis. There are two ways of thinking about such connections: First, we demonstrate why each hypothesis or research question ought to be true—that is, why the concepts ought to be related in the way the hypothesis says they are. This is the theoretical linkage. Second, we show how the concepts are related empirically. This is the operational linkage. Both are necessary if the theory is to be fully elaborated.

Theoretical and operational linkages are also necessary for propositions. The theoretical linkage explains why the proposition should be true, without concern about relations among concepts. The operational linkage shows the type of data that support the proposition.

Likewise, for research questions, theoretical linkages explain the logic of the question and justify asking it. The theoretical linkage may provide hypothesized explanations for varying (and perhaps contradictory) outcomes or answers to the question.

For assumptions, which are not empirically tested, only theoretical linkages are necessary. However, if the assumption is relational, it may be advantageous to specify some elements of the operational linkage.
THEORETICAL LINKAGES

The theoretical linkage gives the theory explanatory power. It explains why the hypotheses, assumptions, and propositions should be true, using at least one of three methods. First, one can cite an existing theory and all of the explanations inherent in the theory. Second, especially if one is working in an area in which theory is not well developed, existing literature can be cited that shows results similar to (or, if one intends to refute a theory, different from) those predicted by the hypothesis. Third, and perhaps most important, researchers must be able to state support for the hypothesis in their own words using their own logic. In fact, it is desirable to state multiple reasons why, logically, the hypothesis should be true. If the researcher cannot state at least one good reason why the hypothesis is true, then the hypothesis is unlikely to receive empirical support. If the researcher can think of 10 good reasons, then the odds are greater that at least one of them will elaborate an empirically supported hypothesis. Of course, it is most desirable for the researcher to use all three methods of explaining the hypothesis under the same reasoning: existing theory, existing literature, and logical reasoning. The more evidence we can muster to support the hypothesis, the more confident we may be that it will be empirically supported.

By these methods, the explanatory power of the theory is increased. Hypotheses, research questions, and propositions are not thrown carelessly into the theory (or tested individually without thought) but are instead incorporated into a whole, showing why the concepts in the theoretical statements ought to behave in the way specified. This forces the researcher to specify ahead of statistical tests at least one good reason why the hypothesis ought to be supported. It therefore reduces the probability that the researcher will thoughtlessly create hypotheses merely for the pleasure of doing data analysis or that hypotheses will be tested only to satisfy minimal curiosity. Theoretical statements ought to be created out of a theoretical whole, and their introduction into a study is best accompanied by a strong theoretical linkage.

In fact, the group of theoretical linkages supporting the hypotheses in a study is itself the theory that the researcher presents. From what other source can it come? The existing theories, literature, and logical statements are the support for the hypotheses, and, in combination, are the theory on which the study rests. Thus, the specification of individual theoretical linkages for each theoretical statement is crucial.

In practice, most research articles are written with the literature review and theory specified in advance of the hypotheses. Unfortunately, once the hypotheses are reached, the reader may not follow the author’s logic in understanding the derivation of the hypotheses, and the theoretical underpinnings of the hypotheses may not be as clear in the reader’s mind as in the author’s. Therefore, it is valuable to summarize the theoretical linkages given previously in one or two paragraphs, immediately following each hypothesis. This will ensure that the hypotheses are in fact directly derived from the literature and theory previously presented and that both author and reader are clear about the theoretical underpinnings of the hypotheses. Table 4.1 shows three brief examples of theoretical linkages. In practice, theoretical linkages should be more complete and elaborated upon.

The examples shown in Table 4.1 illustrate how the three types of theoretical linkage—theory, literature, and logic—can be used together to explain why the hypothesis should be supported. In areas where theory is sparse, or where the scholar is building theory, the specification of existing theory may be missing, or theories tangentially related to the topic may be mentioned. It is perfectly reasonable to build one’s own logical structure to defend hypotheses; this is the creative side of scholarship that should be encouraged.

Although we have been talking about theoretical linkages for hypotheses, propositions, and assumptions should also have theoretical linkages. They are especially important in assumptions, where the reader may or may not agree with the assumption after reading the justification for it.

Creating theoretical linkages for research questions presents a special case. Because no prediction is made by the research question, the theoretical linkage should not take sides—for example, present a local argument for one particular potential answer over another. However, the authors should know enough about the topic to be able to intelligently discuss the possible outcomes. Some literature may predict one outcome and other literature a different outcome. Logical arguments may be made for different outcomes. These sorts of differences should be discussed as the theoretical linkage so that we know the authors are not merely “fishing” for results but rather have thought through the research questions thoroughly and know as much as is knowable about the topic being studied.
The more the media emphasize an issue, the more important people think it is.

IV—Amount of media coverage

DV—Importance of the issue to the public

Operational Definition:

The more the media cover an issue, the more important people believe it is.

Responses: 1 = very unimportant; 2 = unimportant; 3 = important; 4 = very important.

Number of stories about crime

Hypothesis 2:

The more a society is integrated through shared norms, the less likely it is that members will commit suicide.

IV—Level of social integration

DV—Suicide rate

Operational Definition:

The more a society integrates through shared norms, the less likely it is that members will commit suicide.

Responses: 1 = very unimportant; 2 = unimportant; 3 = important; 4 = very important.
### Operational Linkages

Telling how the variables in the hypothesis are related is the job of the operational linkage. Operational linkages may be presented in two forms, visual and statistical. We recommend that researchers, particularly those early in their careers, prepare both sorts of operational linkages. They will take the form of figures and statistical tables.

It is of primary importance that researchers understand that both types of operational linkages must be prepared before any data are collected. If the researchers are able to graphically illustrate what their hypotheses predict and statistically explain how the hypotheses will be tested, then they are more likely to include all variables needed to test the hypotheses. Often researchers who have not clearly thought through their studies collect and analyze their data and then wish that such and such a variable had been included. The graphic and statistical forms of operational definitions make such mistakes highly unlikely.

For beginning scholars who are unfamiliar with statistics, it is at least necessary that they prepare the graphic form of the operational linkage. This requires no knowledge of statistics and helps the researcher think through the potential testing of the hypothesis.

The graphic form of operational linkage is not included in final research projects, which place their emphasis on the degree of statistical support for the hypothesis after data analysis, not on what the hypothesis graphically predicted. On the other hand, research proposals, such as thesis or dissertation proposals, are often aided by the inclusion of graphic operational linkages because they help not only the researcher but the committee in evaluating whether all elements of the study are present and whether the student is ready to proceed with data collection.

### Operational Linkages as Visual Representations

The simplest form of operational linkage is a pictorial representation of the hypothesis. For example, Figure 4.1 shows a simple graphic operational linkage for the hypothesis “The more television violence children view, the more aggressive acts they portray.” The horizontal axis is generally taken as the independent variable and designated as \( x \) (see Figure 4.2), and the vertical axis is assumed to be the dependent variable, called \( y \).
In Figure 4.3, the slanted line is a visual representation of the hypothesis, whereas the dots represent real data collected. Remember, we draw the operational linkage before collecting data. Once we collect data, we will see that the reality of the data do not neatly fit the hypothesized line. The difference between our hypothesized relationship and the data expresses how closely our hypothesis is supported.

When preparing operational linkages as visual representations, we have four considerations: the form of the relationship (linear, curvilinear, or power), the direction in which the variables are related (positive or negative), the coefficients (constant and slope), and the limits (the range within which the hypothesis is supported).²

**Form of the Relationship**

As Figure 4.4 shows, relationships may be hypothesized to be linear, straight lines. In these relationships, one or more units of change in the
**Figure 4.4** Four elements of an operational linkage: the form

*Linear*—As the values of one variable increase, the values of the other variable increase (or decrease).

*Curvilinear*—As the values of one variable increase, the values of the other increase (or decrease) up (or down) to a point and then start off in the other direction.

*Power*—As the values of one variable increase, the values of the other variable increase (or decrease) at an accelerated rate.

---

independent variable are accompanied by one or more units of change in the dependent variable, in a precise and constant manner. Such relationships are frequently hypothesized in the social sciences but are rarely seen as pure cases. Reality rarely fits a straight line; however, most statistics test for the presence or absence of a straight-line relationship. Thus, this basic relationship is commonly used.

Curvilinear relationships may form the letter “U” in an inverted or upright position, or they may be shallow curves in either direction.

Power curves are a special case of curvilinear relationships, where a change in the independent variable has a huge change in the dependent variable at one point in the relationship. There is very little change in the dependent variable initially, then a huge change, then very little change.

---

**Direction of the Relationship**

Here we specify whether the units of a variable change in the same direction (positive) or in opposite directions (negative). In a positive relationship (Figure 4.5), an independent variable’s and a dependent variable’s values change in the same direction (e.g., both higher). In a negative relationship, they change in opposite directions (e.g., the independent variable increases in its value, whereas the values of the dependent variable decrease in value).

---

**Figure 4.5** Four elements of an operational linkage: the direction

*Positive*—As the values of one variable increase, the values of the other variable also increase.

*Negative*—As the values of one variable increase, the values of the other variable decrease.

---

**The Coefficients**

The slope indicates how steep or shallow the line is (Figure 4.6). A steep line indicates that a one-unit change in the independent variable is accompanied by more than one unit change in the dependent variable. A shallow line might indicate that a one-unit change in X is accompanied by less than a whole-unit change in Y.

The constant indicates where the line crosses the y axis. This tells us what the value of Y is for what is generally the minimal value of X. (Exceptions occur where, for example, the scale of the independent variable ranges from negative to positive values, with zero in the center. The constant would probably be where Y crossed the zero point of X.)

---

**The Limits**

As Figure 4.7 shows, increases in X may not indefinitely be accompanied by increases in Y. Take the example of the hypothesis “The more education a person has, the more he or she will read a daily newspaper.”
A person may have indefinite years of education, but there are only 7 days a week available for newspaper reading. In reality, newspaper reading evens off after a relatively small number of years of education, say high school graduation or some college.

The figure shows boundary limits for linear, curvilinear, and power relationships, but the examples illustrate only some of the possible limits that may be hypothesized.

**Operational Linkages as Statistics**

Operational linkages are often expressed in statistical terms, such as a statistically significant and positive Pearson’s correlation coefficient.
variables have the characteristics of being exhaustive (all possible values/categories of the variable are represented) and mutually exclusive (an item can be placed only in one value of the variable). **Ordinal** variables include the characteristics exhaustiveness and mutual exclusivity, as well as order. That is, the numbers assigned to the values must be in a numerical and meaningful order. For example, a variable with *high*, *medium*, and *low* values could have them assigned as 3, 2, and 1, respectively. We cannot logically assign them as 2, 3, and 1, for the order of the numbers would not be the same as the order of the values. **Interval** variables include the characteristics of exhaustiveness, mutual exclusivity, and order, as well as having equal intervals between the numbers and their values. For example, the Celsius temperature scale assigns 100 to the boiling point of water and 0 to its freezing point. These assignments of numbers to these variables are arbitrary once order is achieved, but the scale’s values are equally spaced between 0 and 100. A temperature of 80 is twice as hot as a temperature of 40. The amount of heat gained when moving from a temperature of 10 to 11 is the same as moving from 39 to 40.³ **Ratio** variables have all of the characteristics of nominal, ordinal, and interval variables, as well as having an absolute zero. In the other three levels, zero can be assigned to any value arbitrarily. In ratio variables, zero can only mean *none of the concept*. For example, in the Kelvin temperature scale, zero represents no heat.

**THE WHOLE STORY**

The specification of the operational linkage puts all of the parts of the theory together, and the theory may be shown in brief in Table 4.1. This brings together all of the parts we have been talking about in Chapters 2, 3, and 4. Although none of the three theories mentioned in Table 4.1 is complete, the table does suggest how complete theories may be elaborated. A completely elaborated theory would include all assumptions and hypotheses used, their concepts, and their definitions and linkages. This is an overwhelming task, but one that would help advance social science by making theories much more explicit than they are today and would therefore allow scholars to test the theories and advance them through support, modification, or lack of support. The current state of social science theories is much more vague, with one scholar meaning one thing by a concept and another scholar meaning something else entirely. Once definitions and linkages are specified explicitly, a base will have been established that will permit the growth of theories. Connections among theories may be made more easily, and the state of social science research will advance more quickly and with logical deliberation.

**NOTES**

1. The following is adapted from Hage (1972), pp. 85-110.
2. The terms *constant* and *slope* are familiar to those who know regression statistics, but knowledge of this statistical procedure is not necessary to conceptually understand these terms.
3. There are occasional disagreements over what is interval and what is ordinal. For example, the Likert scale (5 = *strongly agree*, 4 = agree, 3 = neutral, 2 = disagree, 1 = *strongly disagree*) is treated by some social scientists as an interval scale, using the logic that values represent equal-appearing intervals: That is, people interpret the intervals as being equal. Others are more conservative and use this as an ordinal scale.
A frequent occurrence in communication research is to begin a line of inquiry with a hypothesis of broad, across-the-board effects, only to find in subsequent research that the effect is not general but occurs only under certain conditions. What started out as a clean, neat two-variable hypothesis needs to be qualified. The hypothesis is no longer parsimonious; it needs to have qualifying phrases tacked on. And it is no longer as general, and therefore not as potent; it holds up only in certain situations. Meadow (1985) appeared to have been referring to this occurrence when he wrote, “After four decades of exploration, we are left with one answer to the question of media effects—‘it depends’” (p. 158).

Faced with this kind of outcome, we can easily imagine a process with no end. Will further research uncover still more qualifying variables, until it takes knowledge of 50 variables to make a prediction and only a computer can handle the complexity?

Meadow is right, of course, but his conclusion should not be the basis for throwing up our hands in despair. In fact, another point of view is that this is just the place at which things get interesting. As we find out more about the details of these dependencies, we are finding
out something about the complexities of human behavior and the variables that make a difference. The proof of the pudding is that knowledge of the dependencies or the contingent conditions often lets us make more accurate predictions about social interactions and other forms of human behavior. In short, there’s a big difference between saying, “It depends, but I don’t have any idea on what” and “It depends, and the two or three most important variables it depends on are X, Y and Z.”

Part of the solution to this dilemma may lie, as it often does, in taking a middle course. On the one hand, two-variable relationships are probably too simplistic. On the other hand, 50-variable relationships are probably too complicated. Even a path model with just five variables can be so complex that it is difficult to understand. But fortunately there is a range of possibilities in between these. What might make sense as a research strategy in many areas is to explore three-variable relationships, the next level of complexity beyond two-variable relationships. As we shall see, this next step increases the complexity quite a bit. And it seems like a natural step in that it takes us beyond Hage (1972), whose valuable book deals mostly with two-variable relationships.

Further, it may not be necessary to study a huge number of variables all at once. As Hirschi and Selvin (1967) noted, “The increasing success of the life sciences in understanding the human body (surely more highly integrated than the social system) suggests that good research is possible without taking everything into account at once” (p. 22).

This chapter discusses the next step in theory building in the social sciences beyond the formulation and testing of two-variable hypotheses. As Eveland (1997) noted, “Many theories . . . in mass communication and related fields predict more complex effects than the simple linear and additive effect of independent variables” (p. 405).

For instance, the knowledge gap hypothesis suggested by Tichenor, Donohue, and Olien (1970), which has generated a great deal of subsequent research, is basically a three-variable relationship. A graph illustrating the knowledge gap hypothesis is presented in Figure 5.1. The graph shows a relationship between exposure to information and knowledge, with knowledge increasing as exposure to information increases. But it also shows that the rate of increase in knowledge is different depending on the socioeconomic status of the individual. This kind of three-variable relationship is called an interaction. We have an interaction when the relationship between one variable and a second variable is different depending on the values of a third variable.

Hage’s (1972) book focused on two-variable theoretical statements, for which he recommended the form “The greater the X, the greater the Y.” He acknowledged the importance of relationships that are more complicated than two-variable relationships, but he didn’t do very much to deal with them. He wrote:

Through our discussion, we have been concentrating on the problem of interrelating just two variables. In practice, we can and do expect our operational linkages to be more complex than this. The diagrams and forms refer to the effect of X on some Y where X can be a combination of variables. (p. 109)

The idea that X can be a combination of variables oversimplifies the complexity of even a three-variable relationship quite a bit. As we shall see, a set of three variables can be related in five distinctive ways. The situation becomes even more complicated with more than three variables.

Of course, it is a common research strategy to introduce a third variable as a control variable while exploring a relationship. The usual reason for doing this is to test whether a relationship is spurious. This is indeed an important reason for introducing a third variable. Actually, however, there are at least three other important roles that exploring three-variable relationships can play in the development of theory.
ROLES OF THREE-VARIABLE RELATIONSHIPS IN THEORY

One of these important roles is specifying the ways that a hypothesis expressing a relationship between two variables either holds up or does not hold up depending on the state of a third variable. This is sometimes spoken of as specifying the “contingent conditions” (Winter, 1981, p. 236) or as exploring “contingent causation” (Chaffee, 1977, p. 226). As Chaffee noted,

Physically partitioning samples on the basis of contingent orientations that are necessary for a media effect to operate is likely to become more common in future research; this trend is a sign both that we understand quite a bit about the total influence process and that we are going to be able to learn more. (p. 227)

An example of research that specified contingent conditions for communications effects to take place is Hill’s (1985) study of agenda setting by television news. He found that agenda-setting effects are more likely when viewers have prior awareness of news topics through print media exposure or have some college education.

A second major role that exploring three-variable relationships can play in building theory is in clarifying causal relationships by showing the operation of intervening variables. This is the process of analysis that Hyman (1955) called interpretation. Commenting on the area of mass media effects research, Chaffee (1977) stated, “What has been lacking, although it too is beginning to accumulate, is three-variable research in which the psychological processes that intervene between media exposure and its effects are studied” (p. 222).

Chaffee (1977) gave an example of a research area where there is a need for interpretation. He cited the finding that a boy is shown a filmed fight and subsequently acts more aggressively than before. We might conclude that the film has had an effect, but we might not understand the intervening psychological process that led to this effect. In this example, there are several possible intervening processes—the boy may be imitating the behavior he has seen; the boy may be “identifying” with the grown, strong man who was fighting; or the film may have aroused him to do something physically active, and any kind of activity may have reduced that state of arousal. We don’t know which of these possible interpretations is correct until we do some additional three-variable studies.

To take another example, interpretation could be useful in clarifying our understanding of how agenda setting works. We don’t really know much about the process by which the play of news items in the media gets translated into a set of priorities in an individual’s mind. One possibility—and this is only one of many—is that the public has a good understanding of journalism and interprets the cues of news play as indicators of importance. We could test this possibility by introducing as a third variable a new concept called journalistic savvy. This variable would attempt to measure understanding of the various journalism conventions used to indicate story importance. This variable could be measured by having survey respondents indicate their agreement on 5-point Likert scales with statements such as the following:

“If a news story is on the front page of a newspaper, it is important.”

“The most important stories in a television newscast come at the beginning of the newscast.”

“The larger the photograph that accompanies a story, the more important the story.”

Once these kinds of variables have been measured in an agenda-setting study, they can be used as third variables in the kind of three-variable analysis we are recommending.

There are many other good candidates for variables that would help us understand the agenda-setting process by using a third variable to interpret the two-variable agenda-setting relationship. An excellent way to focus one’s theory-building efforts is to think of possible variables of this type, figure out how to measure them, and then conduct a study in which they can be introduced as third variables along with the two principal variables of the agenda-setting process (the media’s agenda and the public’s agenda).

A third important reason for exploring three-variable relationships is that it can help specify the operating component of a global variable (Rosenberg, 1968). For instance, a researcher may find that whether people vote is related to social class. But social class is a global variable with a number of components, including education, income, social prestige, and type of occupation. Which component (or components) of social class is operating to influence voting behavior? Bringing each component variable into a three-variable analysis with social class and voting can help to pinpoint the active variable and in doing so will sharpen our understanding of the causes of voting or not voting.
**FIVE TYPES OF OUTCOMES**

We have been focusing on a general research strategy in which two variables are cross-tabulated and then a third variable is introduced as a control or test variable. This strategy has been referred to as *elaboration* by Paul Lazarsfeld (1955a), who did much of the work to develop it. The elaboration model was first described in a paper that Lazarsfeld presented to the American Sociological Society in 1946. That paper was not published in a journal because “at that time, there was little interest in methodological discussion” (Lazarsfeld, Pasanella, & Rosenberg, 1972, p. 125). The model was first presented in published form by Patricia L. Kendall and Lazarsfeld (1950), and the original Lazarsfeld paper was finally published in 1955 (Lazarsfeld, 1955b). The strategy of elaboration research is discussed in Hyman’s (1955) *Survey Design and Analysis* and Rosenberg’s (1968) *Logic of Survey Analysis*.

Lazarsfeld (1955a) and Hirschi and Selvin (1967) indicated that there are four possible outcomes of this type of analysis. But there appear to be five significant ones.

In the following discussion of this kind of analysis, the terms *independent variable* and *dependent variable* will be used, even though the variables are not necessarily being manipulated in an experiment. *Independent variable* will be used to refer to the variable thought to be the causal variable. *Dependent variable* will be used to refer to the variable thought to be caused, or to be the effect. *Control variable* or *test variable* will be used to refer to the third variable brought into an expanded analysis of a two-variable relationship. An *intervening variable* is “one that is viewed as a consequence of the independent variable and a determinant of the dependent variable” (Rosenberg, 1968, p. 54).

1. **Explanation.** Explanation is one of the possible outcomes of three-variable analysis discussed by Hyman (1972). Explanation occurs when the third variable is causally prior to the independent variable and accounts for the original relationship (see Figures 5.2a, 5.2b, and 5.2c). In this situation, we conclude that the original relationship was spurious, and we discard the original relationship. This is not a total loss, however, because we have learned something about the causal relationship. We have learned that the third variable is a more plausible cause of the dependent variable than the original independent variable.
2. Internal Replication. This term appears to have been introduced by Hirschi and Selvin (1967). Internal replication occurs when the third variable does not affect the original relationship (see Figures 5.3a, 5.3b, and 5.3c). In this case, we conclude that the third variable is not an important variable in this relationship. Our faith in the original two-variable relationship becomes strengthened because we have ruled out one possible alternative hypothesis.

Figure 5.3a  It appears that heavy TV viewers are more likely to give the TV answer to a survey question than light viewers are. But will this relationship hold up once control variables are introduced?

Figure 5.3b  We control for age and find that it has no impact on the original relationship. Heavy TV viewers are more likely to give the TV answer regardless of their age. This is internal replication of the original relationship.

Figure 5.3c  A line graph also shows that the original hypothesis is still supported. The two lines are close together, indicating little or no main effect of age.

3. **Interpretation.** This is another of the outcomes discussed by Hyman (1955). Interpretation takes place when a third variable occurs in sequence between the independent variable and the dependent variable, and the third variable accounts for the original relationship (see Figures 5.4a, 5.4b, and 5.4c). Interpretation and explanation differ only in the time sequence or causal sequence that is attributed to the variables. If the third or control variable occurs prior to the two original variables, we are in a situation in which explanation might occur. If the third or control variable occurs in a time sequence between the two original variables, we are in a situation in which interpretation might occur. The causal sequence is typically something that the researcher must determine. In many cases, it may be very clear which variable occurs first in a causal sequence, but in other cases it may be a matter of debate. Two factors can be used to help us decide whether one variable is causally prior to another: time order and the fixity or alterability of the variables (Rosenberg, 1968). For example, gender is a fixed or unalterable variable, and it is not likely to have been caused by another variable in a data set, such as voting behavior.

---

**Figure 5.4a** It appears that high-education people are more likely to show high political knowledge than are low-education people (data are hypothetical). But what happens when we introduce control variables?

**Figure 5.4b** We control for newspaper reading and find that this variable intervenes causally between education and political knowledge, making the original relationship disappear. This is a case of interpretation—introducing the proper control variable helps us interpret the original relationship (data are hypothetical).

**Figure 5.4c** A line graph also shows that newspaper reading is a better predictor of political knowledge than is level of education. There is a main effect of newspaper reading only, and there is no interaction (data are hypothetical).
4. **Interaction.** This is a term from analysis of variance meaning that the relationship between one variable and another variable is dependent on the value or level of a third variable (see Figures 5.5a, 5.5b, and 5.5c). This is the frequent situation in social science research that elicits the “it depends” conclusion. This outcome helps the researcher to specify the contingent conditions under which a relationship holds up. Hirschi and Selvin (1967) indicate that this is the most common outcome of introducing a third variable. They point out that Hyman (1955) used the term specification for a particularly strong interaction—one in which one partial relationship is stronger than the original relationship, or perhaps opposite in sign.

**Figure 5.5a** It appeared in these data from the 1980s that people higher in education were more likely to favor a nuclear freeze than were people lower in education. Is this a genuine relationship, or is it spurious?

**Figure 5.5b** Introducing political orientation as a third variable shows an interaction—the relationship between the original two variables depends on the value of the third variable. For conservatives, the difference in education makes little difference. For liberals, the difference in education makes a substantial difference.

**Figure 5.5c** A line graph also illustrates the interaction. Line graphs are very useful for detecting and illustrating interactions. If the two lines are not parallel, this indicates that an interaction is present. If the two lines are parallel, there is no interaction.

---

**SOURCE:** Adapted from J. W. Tankard, Jr. (1983, August). Knowledge and opinion on the nuclear freeze: A test of three models. Paper presented to the Communication Theory and Methodology Division, Association for Education in Journalism and Mass Communication, Corvallis, Oregon.
5. **Additivity.** This occurs when both the independent variable and the control variable affect the dependent variable, but the effects are independent and the variables do not interact (see Figures 5.6a, 5.6b, and 5.6c). In the terms of analysis of variance, this is the situation in which you have two main effects but no interaction. This is the outcome that, although quite common, has tended to be overlooked in some discussions of three-variable relationships.

**Figure 5.6a** It appears that heavy viewers of TV are more likely to give the TV answer to a survey question than are light viewers. But what happens when we introduce a control variable?

![Bar chart showing percentage giving the TV answer for light and heavy viewers](chart1.png)


**Figure 5.6b** After we introduce gender as a control variable, the results show additivity. When we control for gender, we see an effect due to television viewing and an effect due to gender. The effects of the two variables “add” up.

![Bar chart showing percentage giving the TV answer for light and heavy viewers by gender](chart2.png)

**PROPER FORM FOR HYPOTHESES**

The recommendation of studying three-variable relationships raises an important question—what is the proper form of hypotheses for three-variable relationships? In Hage’s terms, this is the question of the operational linkage.

Three-variable hypotheses frequently express relationships that are so complicated that they are difficult to put into words. In addition, they are often poorly stated. Eveland (1997) noted that key phrases in statements suggesting interactions (and nonlinearity) include *will depend on, will be different for, should increase the effect, and will occur only for* (p. 405). But, hypotheses using these phrases may not be clear or specific enough.

The statement by Gerbner, Gross, Morgan, and Signorillli (1980) of their “mainstreaming” hypothesis shows how confusing statements of three-variable relationships can be: “By ‘mainstreaming’ we mean the sharing of that commonality among heavy viewers in those demographic groups whose light viewers hold divergent views” (p. 15). This is three-variable hypothesis, with the phrase *demographic groups* covering a number of different possible third variables. But it is difficult to draw in advance the bar chart that would illustrate a result that would fit the hypothesis.
A well-known hypothesis that illustrates the complexity of three-variable relationships is the knowledge gap hypothesis. As stated in the original article by Tichenor et al. (1970), this hypothesis takes the following form:

As the infusion of mass media information into a social system increases, segments of the population with higher socioeconomic status tend to acquire this information at a faster rate than the lower status segments, so that the gap in knowledge between these segments tends to increase rather than decrease. (p. 159)

This statement of the knowledge gap hypothesis is probably about as clear as it can be, but it is noteworthy that it is 49 words long and that a diagram is almost necessary to show just what is meant.

Several forms of hypotheses that we sometimes see presented do not seem to be adequate. For instance, it is not enough to predict that "there will be an interaction." A number of different patterns of interaction are possible, and this kind of statement does not specify which one.

An example given by Hage (1972) also does not seem to suggest a workable general form for a three-variable hypothesis. Hage presented this hypothesis: "The lower the education, when there is little or no income, the lower the power (p. 94)." But this statement leaves out half of the full three-variable table—the half corresponding to the condition of high income.

What seems to be needed in a hypothesis for a three-variable relationship is a very explicit statement of the expected relationship between Variables X and Y for each condition of Variable Z. The hypothesis can almost mirror the form of a table showing a three-variable relationship. For instance, here is an explicit statement of the kind of prediction one might make for a well-known attitude change experiment:

When audience members are high in education, a two-sided message will lead to more attitude change than a one-sided message, but when audience members are low in education, a two-sided message will lead to less attitude change than a one-sided message.

The above hypothesis is also somewhat long, but such length may be necessary to deal with the complexity of a three-variable relationship.

| SOME METHODOLOGICAL CONSIDERATIONS |

Three-variable analysis is a strategy that is applicable with many research methods, including particularly surveys and experiments. It may also be suitable in content analysis research if the object of the content analysis is to test hypotheses about causal relationships.

Various techniques of data presentation and analysis can be used to carry out three-variable analyses. This chapter has stressed visual examination of three-variable relationships with column charts and line graphs. Such visual techniques present some obvious advantages in presenting relationships clearly (Tankard, 1994). Gerbner and his associates in their research on television have used column charts for much of their data presentation (Gerbner & Gross, 1976; Gerbner et al., 1980). Three-variable relationships can also be presented in the form of cross-tabulation tables, and this was the technique used predominantly by such pioneers as Lazarsfeld and Hyman.

Other techniques that can be used to examine three-variable relationships include partial correlation, factorial designs using analysis of variance, and looking at correlation coefficients within sub-groups. An example of the last technique is O'Keefe's (1985) study of the effectiveness of a mass media crime prevention program. This technique has some of the advantages of partial correlation—there are fewer numbers to look at in a table, and the computer can do a lot of the work for you—and some of the advantages of cross-tabulations—you can see which of the five types of three-variable relationships you have.

If a researcher is going to attempt three-variable analysis, one of the important questions becomes which variables to use as control variables. The primary criterion for selection of control variables should be theoretical relevance (Rosenberg, 1968, p. 38). Is the control variable likely to be implicated in the relationship? In most situations, there will be numerous variables that one could select as a third variable, with several purposes in theory building that could be served by selecting each. So the process of selecting control variables should not be a mindless one, and it requires some knowledge of theory.

Finally, a few methodological warnings seem in order for researchers working on three-variable relationships. As more analysis by subgroups is done, replication probably becomes more important, to show that differences between subgroups aren't due merely to sampling
error. Also, there is another danger in looking at differences between subgroups—the danger of post hoc reasoning. One of the problems with post hoc reasoning is that it can mean that nothing is ever falsified. Instead of falsifying a hypothesis, we qualify it. For instance, a researcher might start out with a hypothesis that a certain general effect will occur. In various studies of the hypothesis, however, it might happen that the hypothesis is supported in some studies whereas its logical opposite is supported in other studies. The tendency in these cases seems to be not to conclude that there is evidence contrary to the original hypothesis but to revise it to state that the outcome is dependent on a third variable. Although this procedure is essentially the approach being presented in this chapter, it does raise some questions. Falsification, as used by Popper (1959), and strong inference, as used by Platt (1964), do not seem possible with this approach.

CONCLUSION

The main purpose of this chapter is to extend our thinking in theory building beyond two-variable relationships to three-variable relationships. The chapter points out that introducing third variables into an analysis can serve several important purposes in theory building beyond just checking for a spurious relationship.

Much of social science theory seems to be moving toward a shift from statements of very general relationships between two variables to more qualified statements of when effects are likely to occur. In general, we’re moving away from strong effects models of human behavior and toward contingent effects models. The conclusions we are reaching seem to be similar to the conclusion Roger Brown (1958) drew some years ago about the effectiveness of the propaganda devices—that they are “contingently rather than invariably effective” (p. 306).

Three-variable analysis is highly suitable for this kind of theory development. It seems to be the next logical step beyond the two-variable approach.

In this chapter, we will cover why the theory builder might want to use more than three variables in a hypothesis and how to construct theoretical statements relating four or more variables. We will learn how to extend the three-variable strategy to these more complex systems, how to express such multivariate relationships in hypothesis form, and how to think about theoretical and operational linkages for the multivariate relationships of complex systems.

In building theory, the importance of well-expressed theoretical statements cannot be overemphasized. If formulated properly, theoretical statements help the theory builder stay focused on the task at hand. An added benefit is that clear theoretical statements help communicate to others what one is attempting to do. As we have seen, even the study of how one concept is related to another concept can be sidetracked by a poorly conceived theoretical statement. As we also have seen, the addition of just one more concept to our theory will significantly increase the complexity of our work, making it all the more important.