THEN: Organization structure is operationally defined as the number of rules, degree of rules followed, task discretion . . . (and variables of other constructs of the concept of organization structure).

The *construct validity* of this set of deductive conditional propositions is established by showing that each consequent follows its antecedent by the very definition of the antecedent. The consequent is true if the antecedent is true.

As this example illustrates, many other constructs and variables could be used to operationally define Weber's concept of structure. Moreover, constitutive definitions of other theories of organization structure would result in many other operational variables. The point is that the variables included in any research model represent a sampling of many possible observable variables of a construct both from a given theory and from different theories. That being the case, it is crucial to select only those operational variables of the theory that are most important or relevant for addressing the research question being investigated.

This principle also applies to sampling hypotheses from propositions because the principles of traveling the ladder of abstraction for concepts also apply to propositions. Chapter 4 noted that a theory consists of propositions and hypotheses that differ by levels of abstraction: propositions are relationships among theoretical constructs, while hypotheses are relationships among observable variables. Hypotheses are sampled deductions from one or a few basic propositions using causal conditional propositions. Following Osigweh's (1989: 585) maxims, we climb the abstraction ladder by extending the conceptual breadth of hypotheses into more general propositions, while reducing their connotation (thereby increasing simplicity). As we climb, we rise to fewer and more general propositions as we move from conclusions (hypotheses) to the premises that entail them (propositions and assumptions) (Kaplan 1964: 298).

Conceivably, an infinite number of hypotheses can be derived from a theoretical proposition (Dubin 1976). However, any research model includes only a small sample of observable hypotheses that can be deduced from, or induced to, a theoretical proposition. That being the case, Stinchcombe (1968*b*) and Giere (1999), among others recommend a sampling strategy of selecting only those hypotheses that represent diverse tests of a proposition. The greater the number of divergent hypotheses that do not reject a proposition, the more credible the proposition. A second way to increase the plausibility of a theoretical proposition is to rule out hypotheses derived from rival alternative propositions. At a minimum, to be viewed as credible, hypotheses derived from a new proposition should provide better explanations of a phenomenon than hypotheses reflecting the status quo explanation. Stinchcombe (1968*b*) discusses how this basic inductive process of science should lead researchers to design *crucial experiments* where evidence in support of one theoretical proposition implies the rejection or negation of a rival alternative proposition.

Over the course of a study, researchers often change their initial understandings of the conceptual domain being studied. Theoretical sampling is seldom static; it is an ongoing sensemaking process (Weick 2005). Shadish et al. (2002: 21) note the following:

> There is a subtle interplay over time among the original categories the researcher intended to represent, the study as it was actually conducted, the study results, and subsequent interpretations. This interplay can change the researcher's thinking about what the study particulars actually achieved at a more conceptual level, as can feedback from readers. But whatever reconceptualizations occur, the first problem of causal generalization is always the same: How can we generalize from a sample of instances and data patterns associated with them to the particular constructs they represent? (Shadish et al. 2002: 21)

POPULATION SAMPLING

In addition to drawing inferences to theory, researchers also want to generalize study findings to various people (units), settings, and outcomes in a population. If the population of interest is known, then it is possible to identify the variations in units, settings, and outcomes that exist in that population. The basic sampling strategy is to ensure that the range of variation in the target population is adequately represented in the study's sample of observations. Singleton and Straits (2005: chap. 5) provide a useful review of the major steps in sampling design, which is a plan of how cases are to be selected for observation. This plan involves three major steps: (1) define the population; (2) construct the sampling frame; and (3) implement a probability or nonprobability sampling strategy.

The first step is to identify the target population, which is the particular collection of units and settings to which a researcher would like to generalize study findings. Singleton and Straits (2005: 113) credit the sociologist, Kenneth Bailey (1982), for noting an important distinction between experienced and novice researchers in how they approach sampling:

> The experienced researcher always gets a clear picture of the population before selecting the sample, thus starting from the top (population) and working down (to the sample). In contrast, novice researchers often work from the bottom up. Rather than making explicit the population they wish to study, they select a predetermined number of conveniently available cases and assume that the sample corresponds to the population of interest. Consider a sample consisting of 'randomly' chosen passersbys at a shopping center on a Saturday afternoon. What could this sample possibly represent? There is simply no way of knowing until an intended or target population is defined. (Singleton and Straits 2005: 113)

Defining a target population depends upon the unit of analysis and the research question, and involves specifying the cases that are to be included and excluded in the population. If the unit of analysis is individuals, some combination of exposure to the independent variables or treatments being investigated and the selected demographic characteristics such as age, gender, race, and education is typically used. For example, in a study of how healthcare physicians and managers solve problems, Schultz (2001) stratified his target population into individuals who obtained a medical degree (MD) or managerial degree (MBA or MHA) and who varied in age, gender, and years of experience working in supervisory positions of managed healthcare systems. In this experiment Schultz randomly assigned individuals from the two strata of educational degrees in the target population to two problem solving tasks, and then compared their results by statistically controlling for their age, gender, and years of working experience.

As this example suggests, defining the target population is closely related to constructing the sampling frame, which identifies the set of all cases from which the sample is actually selected (Singleton and Straits 2005: 116). This is the second step and can be done by either listing all the cases in the population or by developing a rule that defines membership in the population. Oftentimes it is not possible to identify all members of a target population. A census listing of all members of a target population may not exist. Instead, researchers often rely on a rule stipulating criteria for inclusion and exclusion in the target population. For example, Schultz developed and used a rule that all members of his target population must be working in a supervisory role in a managed healthcare system; anyone not satisfying these conditions was excluded from his target domain. However, his sampling frame was ambiguous with respect to the geographical location of the target population. As a result, the rule that Schultz used to specify his sampling frame did not provide a geographical basis for sampling individuals from different regions in the US and other countries that are generally known to have different healthcare cultures and practices.

The third major step in sampling is to select cases from the target population as defined by the sampling frame developed in step two. Singleton and Straits (2005) discuss two general procedures—probability and nonprobability sampling—that are typically used to select a sample that is representative of a target population in a study. Probability sampling includes simple random sampling, stratified random sampling, and cluster sampling.

- *Simple random sampling* consists of a random selection from the entire population that makes it equally possible to draw any combination of cases from the target population. Using a table of random numbers to select cases from a population, for example, random sampling has the scientific advantage of applying the principles of probability sampling theory to calculate sampling error and estimate sample precision.

- In *stratified random sampling*, the population is divided into strata and independent random samples are drawn from each stratum. This strategy is particularly appropriate for selecting and comparing naturally occurring events or treatments in a target population that are difficult to manipulate experimentally. For example, comparisons between married and divorced couples on child rearing behaviors may only be possible by stratifying the target population into married and divorced couples and then randomly selecting couples from each stratum to examine their parental behaviors.
- *Cluster sampling* is often used when it is impossible or impractical to list all members of a target population. In cluster sampling, the population is broken down into groups of cases, called clusters, which consist of natural groups, such as geographical states, regions and cities, or types of organizations, such as colleges, churches, and businesses.

Nonprobability sampling refers to the non-random selection of cases for a study. Singleton and Straits (2005) discuss a variety of nonprobability sampling procedures, including convenience, purposive, and quota sampling. Since nonprobability samples are not randomly selected, they have two weaknesses: no control for investigator biases in selecting units and not being able to predict variations among sampled units based on probability sampling theory.

SAMPLE SIZE

A final sampling decision is determining the appropriate number of cases to sample in a study. Sample size considerations include: (1) the heterogeneity of the populations; (2) the desired precision in determining magnitudes of effects; (3) the type of sampling design; (4) the availability of resources; and (5) the number of breakdowns planned in data analysis (Singleton and Straits 2005: 140). A discussion of the mathematical statistical analysis for determining the power of tests to achieve statistical significance is beyond the scope of this chapter. Statistical textbooks and web sites are widely accessible for calculating the size of sample required in a study to estimate the power of significance tests for various statistical models.[1]

One consideration that is often overlooked in determining sample size is equating statistical significance with practical significance of a test. Walster and

[1] See, for example the following web sites—*Supercourse—Survey sample size* from the University of Pittsburgh (at: http://www.lib.umn.edu/libdata/link.phtml?page_id=1187&element_id=34881); *Statistical considerations for clinical trials* from Harvard University (at: http://www.lib.umn.edu/libdata/link.phtml?page_id=1187&element_id=34882); *Statistics Calculator* and *Power Calculator* from UCLA (at: http://www.lib.umn.edu/libdata/link.phtml?page_id=1187&element_id=34884); *The Survey System* from Creative Research Systems (at: http://www.lib.umn.edu/libdata/link.phtml?page_id=1187&element_id=34885).

Cleary (1970) pointed out that classical hypothesis testing methods do not necessarily lead researchers to make rational decisions. They argue that the problem is not with classical methodology, but with the way it is used. Conventional procedures for determining the power and statistical significance of a test are often not consistent with the practical significance of a test. It is well known that a researcher can control the power of a statistical significance test by manipulating sample size. But this form of statistical significance should not be confused with practical significance. The latter reflects a judgment by users of study findings on what magnitudes of effect and levels of probability they consider trivial versus those large enough to convince them of altering their behavior with respect to the test in question. For research findings to be relevant to users, Walster and Cleary (1970) advise researchers to select a sample size that equates this qualitative notion of practical significance with statistical significance.

Measurement and Frames of Reference

Once a set of variables has been selected to represent the constructs of interest in a research model, then attention can turn to measuring these variables. Measurement is the process of assigning numbers or labels to variables of units in order to represent their conceptual properties (Singleton and Straits 2005: 76). Fundamentally, measurement represents a problem of conceptualization. Typically, it begins by descending the ladder of abstraction to recast theoretical constructs into observable variables, and select procedures and indicators to measure these variables in ways that are reliable (i.e., replicable) and valid (i.e., capture their intended meanings).

In the physical sciences variables are typically measured with standardized instruments, for example, to find the temperature, mass, density, and force of material objects. In contrast, social scientists examine individual and collective properties that often cannot be observed directly, are too complex for any one person to observe, and for which no uniform or standardized measures exist. As noted in the second section of this chapter on units of analysis, many individual attitudes and behaviors (such as job satisfaction and learning) are based on psychological constructs that cannot be observed directly; they require individuals to express their subjective perceptions and attitudes through the use of questionnaires and interviews. In addition, many properties of collective units of analysis are too complex for one to observe. Organizations, for example, typically consist of many people, groups, and levels with diverse goals, structures, and activities. Measuring these collective properties must often rely on informants, such as top or middle-level managers. However, research demonstrates that perceptions of a few managers are often not generalizable to the entire organization (Dearborn and Simon 1958; Porter 1958).

The frames of reference that individuals take in answering questions dramatically influence their judgments (Guilford 1954; Smith et al. 1969). When two persons with different frames of reference are exposed to the same situation or stimulus, they select different aspects as pertinent to their judgments and provide different summary evaluations of that situation. Frames of reference are the internal standards or cognitive filters a person uses in describing or evaluating a situation (Helson 1964). As applied to measurement, it is useful to examine at least two interlocking issues that influence a respondent's frame of reference: (1) the immediate characteristics of the stimulus or situation to which a person is exposed; and (2) the systematic and unsystematic ways in which individuals respond to the stimulus or situation as a result of prior experiences, dispositions, and roles.

The first issue requires an examination of how a respondent's frame of reference is influenced by the composition of the measurement instrument itself and the setting in which respondents complete it. Specifically, as the top of Table 6.2 outlines, the nature, complexity, referent, and time perspective of

Table 6.2. Development and evaluation of a measurement instrument

Frames of reference in developing a measurement instrument

- Perceptual selectivity in determining human judgments is dramatic. A frame of reference is the cognitive filter a person uses to respond to questions:
    1. Time perspective of questions.
    2. Behavioral, cognitive, or emotional phenomena.
    3. Descriptive or evaluative measures.
    4. Number of intervals or points on answer scale.
    5. Anchors or cues on answer scales.
    6. Unit of analysis.
    7. Respondent or informant role.

Evaluating a measurement instrument

Intrinsic validity—do the measures capture the intended constructs?

- Reliability estimates
    - Repeated, parallel, split half, & multiple measurements
    - Coefficient alpha and the number of items in index
    - Breadth of construct being measured
- Convergent & discriminant validity
    - Factor analyses of all items from several indices
    - Multi-trait, multi-method matrix
    - Median correlations with other items
    - Parallel measures

Extrinsic validity—what are the measures in the instrument good for?

- Conform to theory
- Discriminate different types of units
- Predict or explain criterion/outcome
    - Concurrent validity
    - Predictive validity

Source: Van de Ven and Ferry (1980).

questions and the anchor points on an answer scale have been found to significantly influence a respondent's frame of reference at the point of measurement (Smith et al. 1969; Van de Ven and Ferry 1980). To the extent that a measurement instrument takes these factors into account explicitly, a researcher can control one of the major sources of variation in respondents' frames of reference and thereby have a better understanding of the judgments made by people about the individual and the organizational phenomena of interest. Van de Ven and Ferry (1980: 57–74) provide a useful discussion of the key factors to consider in structuring the frame of reference of questions in a questionnaire or interview instrument.

In addition to the effects due to composition and administration of the measurement instrument itself, there are systematic and unsystematic effects on frame of reference due to the position, past experiences, and predilections of the respondent. The systematic effects include those individual differences in respondents that are known, as a result of previous theorizing or research, to influence respondents' judgments in predictable ways. For example, judgments about individual and group behavior in organizations have been found to differ systematically when respondents occupy different positions and levels in an organization (Porter 1958; Ghiselli 1973; Bouchard 1976). These systematic differences are addressed by developing and implementing a data collection plan that samples respondents or informants from diverse organizational positions and roles. Depending on the variables that are measured, the responses of multiple and diverse informants are then compared and averaged to obtain aggregate scores of organizational groups. Unless there are good reasons to believe that the judgments of one particular informant group are more important or accurate than another, the responses of different informant groups are typically weighted equally in the aggregate collective score (Van de Ven and Ferry 1980).

In studies of job satisfaction and other attitudinal characteristics of organizations (e.g., climate and morale), perceptions have been found to differ systematically among respondents of different age, gender, education, social background, and job tenure in the organization (Smith et al. 1969; Dunnette 1976). These individual difference factors are commonly used as stratification variables when reporting norms for instruments measuring various attitudinal dimensions of jobs and organizations. One reason for this strategy is to statistically control for differing frames of reference of respondents when evaluating a measurement instrument. When these individual difference factors are not explicitly included in a research model, they are often measured and treated as extraneous variables (as discussed before).

The unsystematic effects on frame of reference include a host of unknown predilections, personality orientations, and contextual factors within respondents that influence their individual judgment of a given stimulus in different ways. For example, a sickness in the family, a recent extremely happy or sad incident, and the psychological mood of a respondent at the time of data collection undoubtedly influences his/her answers to questions (Guilford 1954). However, these kinds of influences on frames of reference are unsystematic in the sense that they are expected to be randomly and normally distributed among the sample of respondents or informants and will therefore cancel out statistically when judgments are averaged together. These kinds of unsystematic disturbances on frames of reference are the basis of the argument for obtaining the perceptions of many judges or informants to measure various organizational phenomenon. Classical test theory demonstrates that reliability of a measure increases by increasing the number of judges (Lord and Novick 1968).

Many additional tasks, beyond the scope of this chapter, are involved in developing and evaluating measurement instruments and procedures. Some of the tasks involved in evaluating a measurement instrument are outlined at the bottom of Table 6.3. Readers are referred to Singleton and Straits (2005) and Van de Ven and Ferry (1980) for useful discussions and examples of the process of measurement in social research, and procedures for evaluating various indicators of the reliability and validity of a measurement instrument.

Data Analysis

Scientific inquiry involves a repetitive interplay between theoretical ideas and empirical evidence. Data analysis takes place whenever a research model and data are compared. This comparison occurs whenever the researcher struggles to bring order to, or to make sense of, his/her observations. I suggest below that an engaged scholar should not struggle alone; much help is available if he/she involves other research colleagues, users, and practitioners.

Different methods of data analysis are appropriate for different variance research models. There is no need to enumerate them here since several excellent sources are available for guidance. Yin (2003) and Miles and Huberman (1994) provide useful ways to tabulate, display, and analyze case study data obtained from documents, archival records, interviews, direct observations, participant observations, and physical artifacts. Singleton and Straits (2005) emphasize survey research methods, and discuss methods to edit, code, enter, clean, and document survey data in computer files before steps are undertaken to analyze the data. Pedhazur and Schmelkin (1991) and Neter et al. (2005) provide detailed discussions with examples and software of descriptive and inferential statistics for analyzing multivariate causal models with survey data. Finally, Shadish et al. (2002) focus on analyzing data in order to draw generalized causal inferences from experimental and quasi-experimental designs.

Whatever the data analysis methods and research models that are compared, one thing is clear. A single pass in making sense of data and models is seldom sufficient. Numerous iterations are typically required, and this process is greatly facilitated by engaging others in each iteration. I have found it useful to begin by conducting a preliminary analysis of the research question, model, and data analysis, and then conducting two workshops—one with research colleagues and one with key users or practitioners from the organizations in which the research is being conducted. The workshop with research colleagues tends to provide very useful feedback for refining the analytical aspects of the research model, data analysis procedures, and situating the findings in the research literature. Review sessions with users and practitioners often generate a different kind of feedback dealing with the potential applications and implications of study findings, as well as exploring ways to modify the model and data to examine the research question in more penetrating or relevant ways. Sometimes this includes further data collection that host organizations are often happy to provide (since they raised the further research question). I typically conclude these workshops by indicating that the research team will investigate the most plausible suggestions made and schedule another review session to share the findings. I also ask for volunteers who are willing to help or advise the research team to undertake the next iteration of analyzing study data.

And so the next iteration of the process unfolds, culminating with a second round of workshops in which the revised study findings are presented in a report and discussed. Feedback from the second workshops are typically useful for refining the study report, and for concluding the study. In several instances my research team was invited to continue or expand the research into a longitudinal study with the support and collaboration of research colleagues, users, and practitioners.

I have learned several lessons from conducting these research workshops. First, inviting feedback on research findings can easily lead to 'scope creep' of the research agenda into unexpected and distracting directions. Being clear about your research question and agenda are critical for being open to suggestions and negotiating them in ways that add value and direction to the research objectives. Second, some of my greatest insights and learning experiences about research questions have come from these research workshops with colleagues, users, and practitioners. These learning insights would not have been gained had my research team not involved others in analyzing and reporting the findings. Among the insights were learning different ways to interpret and construct study findings, and understanding the threats to the validity of a study (discussed next) and how these threats might be ameliorated. Although these engaged scholarship principles of involving others in data analysis and interpretation entail much more work than 'going it alone,' the insights and learning gained from engaging a community of research colleagues, users, and practitioners are 'priceless.'

Validity

Shadish et al. (2002: 34) define validity as the approximate truth of an inference or knowledge claim of a causal relationship based on evidence that supports that inference as being true or correct. They ground their concept of validity in a correspondence theory that says that a claim is true if it corresponds to the observed world. Philosophers have argued that correspondence theory is compromised because the data to which a claim is compared are themselves theory-laden and so cannot provide a theory-free test of that claim (Kuhn 1962). While recognizing that correspondence theory is vulnerable to this criticism, they point out that among variance researchers this correspondence theory is 'the nearly universal scientific concern of gathering data to assess how well knowledge claims match the world. Scientists also judge how well a given knowledge claim coheres with other knowledge claims built into accepted current theories and past findings' (Shadish et al. 2002: 35).

Over the years Campbell and his colleagues (Campbell 1957; Campbell and Stanley 1963; Cook and Campbell 1979; Shadish et al. 2002) have developed a validity typology that has been widely adopted by social scientists. The typology consists of four related criteria for assessing four kinds of inferences typically drawn about causal inference from an experimental study: statistical conclusion validity, internal validity, construct validity, and external validity. Shadish et al. (2002: 38) define these criteria as follows.

1. *Statistical conclusion validity* refers to the appropriate use of statistics to infer whether the presumed independent and dependent variables covary.
2. *Internal validity* refers to whether their covariation resulted from a causal relationship.
3. *Construct validity* refers to whether inferences can be generalized to higher order constructs that represent sampling particulars in a study.
4. *External validity* refers to whether inferences of causal relationships hold over variations in persons, settings, treatment, and measurement variables.

Thus, while internal and statistical conclusion validity focus on whether a cause-and-effect relationship is evident in a particular study, construct and external validity refer to generalizations of the study to theory and populations of interest, respectively. These four criteria for evaluating the

Table 6.3. Threats to validity of experimental results

Internal validity: Is the relationship causal or would the relationship exist in the absence of any treatment or variation in the independent variable? Checklist of threats:
1. History.
2. Maturation.
3. Instrumentation.
4. Testing.
5. Statistical regression.
6. Selection.
7. Mortality (attrition).
8. Ambiguity about direction of causation.
9. Contaminations equalizing groups.

Statistical conclusion validity: Are the results due to chance? Possible threats:
1. Statistical power: sampling the wrong number of observations where statistical significance does not equal practical significance.
2. Fishing expedition: maximizing on chance with numerous statistical tests.
3. Reliability of measures.
4. Reliability of treatments—lack of standardization of procedures.
5. Random irrelevancies in experimental settings.
6. Random heterogeneity of respondents.

Construct validity: Do the model findings generalize to the theory? Possible threats:
1. Invalid constitutive definitions of theoretical and empirical terms.
2. Mono-method bias—use of only one procedure to measure variables.
3. Hypothesis guessing—participants guess the hypothesis.
4. Evaluation apprehension—participants present positive impression.
5. Experimenter expectancies that bias the data.
6. Confounding levels of constructs.
7. Interaction of different treatments.
8. Interaction of testing and treatment (especially with pre-testing).
9. Restricted generalizability across constructs.

External validity: Do the findings generalize to the intended population? Possible threats:
1. Not knowing what treatment caused the effect when multiple treatments are used.
2. Did pre-test affect treatment that limits inferences beyond experiment participants?
3. Inferring results beyond the pool of selectively recruited participants?
4. Inferring results to other settings or organizations than those examined?
5. Inferring treatment results to different historical settings.
6. Unrecognized side effects of treatment.

Sources: Campbell and Stanley (1963); Cook and Campbell (1979); and Shadish et al. (2002).

quality of experimental findings are discussed in detail by Shadish et al. (2002; chaps. 2 and 3). Table 6.3 provides a summary reference of the criteria often used to assess threats to these four kinds of validity in experimental studies.

Conclusion

This chapter reviewed some of the basic issues, decisions, and suggestions in designing a variance research model. A variance research model represents the theory as a causal conditional relationship among variables of units that are sampled, measured, and analyzed in accordance with experimental design procedures. I discussed eight key issues that are normally addressed in designing a variance research study. They are the following:

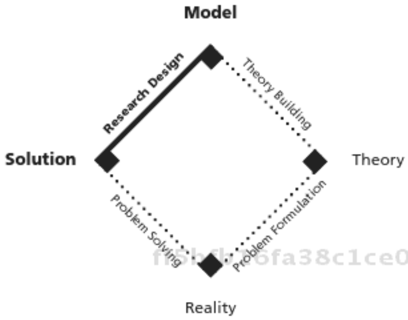
1. Any study reflects the perspectives of certain stakeholders and assumes much tacit knowledge of the particular research context or setting. Researchers are often not aware of the values and assumptions underlying their scientific practices. They become apparent by involving key stakeholders in developing the research question, identifying the key variables and relationships in a research model, and in designing the research. Involving key stakeholders in these issues increases the likelihood that study findings capture the perspectives and tacit knowledge embedded in a research question and model being examined.
2. A research study should clearly identify the unit of analysis, which refers to the entities or objects being studied. Typically, the research question stipulates the entities or objects being examined. However, the units of analysis may not be so simple when examining social collectives existing in a nested hierarchy of individuals in groups or organizations and more encompassing collectives. In these cases the unit of analysis may not be the unit of observation, and special precautions should be taken to avoid individualistic and ecological fallacies.
3. A variance research model consists of one or more causal conditional 'if–then' propositions that are assumed to hold in specified conditions. Most social scientists adopt a probabilistic or manipulative view of causation, and rely on covariation, temporal precedence, and the absence of spurious factors to indicate causation between independent and dependent variables. To deal with plausible spurious or extraneous factors, researchers often add a number of control, moderating, and mediating variables to their causal model. This makes the model more complicated and difficult to examine empirically. In general, parsimonious models are preferred to complex ones due to an ease in understanding and empirical examination. When multiple causal relationships are at play, researchers might include only those factors that are proximate and controllable from the perspective of the key study units or users.
4. Causal models can be examined with a wide variety of randomized, quasi-, and non-experimental designs. Although evidence of causation is strongest with randomized experimental designs, random assignment, and manipulation of treatments are often not possible in a given study. Pragmatic constraints often require researchers to adopt less-than-ideal designs for addressing research questions. That being the case, it is important to assess how any study is vulnerable to threats of validity (see point 8 below), and explore ways for dealing with them.
5. In most social science studies there is tension between the local and particular nature of a research study and the general inferences researchers would

like to draw from their studies. These tensions are addressed by sampling the units, constructs, observations, and settings that are examined in the research. These sampling decisions should be guided by the construct validity of generalizations to an intended theory and the external validity of generalizations to a target population. Furthermore, the sample size or number of cases observed should be chosen to equate the statistical significance of a test with its practical significance to key stakeholders or users of a study. It makes little sense for a study to produce statistically significant findings that are considered trivial by key stakeholders.

6. Measurement is fundamentally a conceptual problem. Typically, the first step in measurement requires defining theoretical constructs into observable variables, and then selecting procedures and indicators to measure these variables in reliable and valid ways. Frames of reference dramatically influence how individuals answer questions and provide data to a researcher. As noted in Chapter 1, social science researchers only obtain the information that organizational participants or respondents are willing to provide. The more a researcher is aware of respondents' frames of reference, the better the measures and their interpretations. As Table 6.2 outlined, the composition of questions in a measurement instrument largely determines how responses are to be interpreted. In addition, a variety of individual difference factors (age, gender, role, experience, personality) influence frames of reference in systematic ways. These systematic factors are typically measured and controlled statistically to examine causal relationships. Finally, the unsystematic effects on frames of reference are assumed to cancel out statistically if the size and distribution of responses in a sample reflects a normal distribution.
7. Data analysis occurs whenever a research model and data are compared. Many research methodology texts provide extensive methods and statistical programs for analyzing data to examine different variance research models. A key suggestion in this chapter is that researchers use the techniques that fit the research question and model. To guide this process, I suggested that researchers conduct workshops with research colleagues and key users or practitioners to obtain feedback on preliminary study findings.
8. No study is perfect. To varying degrees, each is subject to some combination of threats to internal, construct, external, and statistical validities. Assessing the design of a study in terms of these criteria (as listed in Table 6.2) provides a useful checklist of the strengths and weaknesses of a study.

Research design is typically viewed to be a technical project undertaken by researchers trained in experimental research design and statistics. Understanding the technical considerations in experimental design, sampling, measurement, statistical analysis, and inference are crucial to scientific inquiry. Hopefully this chapter has shown numerous instances in which these technical decisions need to be informed by the interests and perspectives of the key stakeholders of a study. Principles of engaged scholarship—identifying, involving, and negotiating the perspectives of key stakeholders—are not only necessary for understanding the purposes and interests served by a study, but also for incorporating their values, interests, and tacit knowledge into the design of a study. Researchers who engage others in the design and conduct of their studies are more likely to develop research findings that penetrate more deeply and have greater impact for theory and practice about the research question being examined than those researchers who 'go it alone.'

7 Designing Process Studies



There is a growing interest in understanding processes of change and development in individuals, groups, organizations, and other social entities. Process studies are undertaken to examine research questions dealing with how things change and develop over time. Chapter 5 reviewed the philosophical assumptions underlying process research and how they differ from variance models. This chapter discusses some of the operational issues and decisions involved in designing process models to either develop or test a process theory. These issues, outlined in Table 7.1, include: clarifying the meanings and theories of process, designing field studies to address process questions, observing and collecting data about process events over time, and analyzing these data into coherent and useful process theories.[1] By necessity these issues are discussed in sequential order. In practice they are highly interdependent and need to be treated in an iterative manner. Poole et al. (2000) provide a more detailed book-length treatment of these issues.

Following a discussion of the process research design issues listed in Table 7.1, this chapter presents an example of designing a study to evaluate an influential process model of organizational growth developed by Greiner (1972). The example also illustrates how valuable insights and learning can be gained by engaging in conversations with others when designing research—in this case between the process theorist (Prof. Larry E. Greiner) and modeler (me).

The chapter concludes on a motivational note addressing concerns often expressed by junior faculty and doctoral students about the amounts of time, resources, and contacts needed to conduct longitudinal process studies.

[1] This discussion does not exhaust the issues that confront process researchers, but in my experience it covers most of the critical choices in designing field process studies of organizational innovation and change. Other good sources for designing longitudinal organizational studies include Galtung (1967), Huber and Van de Ven (1995), Kimberly and Miles (1980), and Miller and Friesen (1982).

Table 7.1. Key issues, decisions, and suggestions for process research in field studies

Issues	Decisions	Suggestions
Formulating the process research plan		
1. Meaning of process	A category of concepts or a developmental sequence?	Process research is geared to studying 'how' questions.
2. Theories of process	Examine one or more models?	Apply and compare plausible alternative models.
3. Reflexivity	Whose viewpoint is featured?	Observe change process from a specific participant's viewpoint.
4. Mode of inquiry	Deductive, inductive or retroductive?	Iterate between deduction and retroduction.
5. Observational method	Real-time or historical observations?	Observe before outcomes are known.
6. Source of change	Age, cohort or transient sources?	Develop parallel, synchronic, and diachronic research design.
7. Sample diversity	Homogeneous or heterogeneous?	Compare the broadest range possible.
8. Sample size	Number of events and cases?	Focus on number of temporal intervals and granularity of events.
9. Process research designs	What data analysis methods to use?	Match data analysis methods to number of cases and events.
Measuring & analyzing process data		
1. Process concepts	What concepts or issues will you look at?	Begin with sensitizing concepts and revise with field observations.
2. Incidents & events	What activities or incidents are indicators of what events?	Incidents are observations; events are unobserved constructs.
3. Specifying an incident	What is the qualitative datum?	Develop decision rules to bracket or code observations.
4. Measuring an incident	What is a valid incident?	Ask informants to interpret and verify incidents.
5. Identifying events	What strategies are available to tabulate and organize field data?	Apply a mix of qualitative and quantitative data analysis methods.
6. Developing process theory	How to move from surface observations to a process theory?	Identify five characteristics of narrative theory.

Process questions of how things change and develop over time require longitudinal data that can be obtained either from historical archival files or from a real-time field study of a change process. Whether the data are obtained from archival sources or from field studies, I advise researchers not to go it alone; instead, they should engage and collaborate with other scholars (typically senior colleagues) who are conducting process studies or have access to longitudinal process data.

Formulating the Research Plan

CLARIFY MEANINGS OF PROCESS

Process studies are centrally concerned with how change unfolds in the entities or things being studied. This chapter focuses on organizational change to exemplify methods for designing process studies. *Organizational change is defined as a difference in form, quality, or state over time in an organizational entity* (Van de Ven and Poole 1995). The entity may be an individual's job, a work group, an organizational subunit, strategy, or product, the overall organization, or a community or population of organizations. Change can be empirically determined by longitudinal observations of the entity over two or more points in time on a set of dimensions, and then noticing a difference over time in these dimensions. If there is a noticeable difference we can say that the entity has changed. Much of the voluminous literature on organizational change focuses on the nature of this difference, and the processes that explain how it unfolds.

Two different definitions of 'process' are often used to explain change: (1) a category of concepts or variables that pertain to actions and activities; and (2) a narrative describing how things develop and change (Van de Ven 1992). As discussed in Chapter 5, when the first definition is used, process is typically associated with a 'variance theory' methodology (Mohr 1982), where an outcome-driven explanation examines the degrees to which a set of independent variables statistically explain variations in some outcome criteria (dependent variables). The second meaning of process takes an event-driven approach that is often associated with a 'process theory' explanation of the temporal order and sequence of change events based on a story or narrative (Abbott 1988; Pentland 1999; Poole et al. 2000; Tsoukas 2005). These two definitions represent very different views of process, and the definition that researchers adopt influences the questions they ask, the research methods they employ, and the contributions they make. Hence, at the outset of a study, it is important to clarify the meanings of process.

Process as a Category of Concepts

Studies of process in the social sciences typically treat process as a category of concepts of individual and organizational actions, such as communication frequency, work flows, decision-making techniques, as well as strategy formulation, implementation, and corporate venturing. In this usage, process refers to a category of concepts that can be distinguished from other categories of concepts, such as organizational environment, structure, and performance. Like these other categories, process concepts are operationalized as variables and measured as fixed entities (variables), the attributes of which can vary along numerical scales from low to high. Studies that adopt this definition of process typically examine research questions dealing with the antecedents or consequences of change. As discussed in Chapters 5 and 6, these kinds of questions call for a variance research design of the causal factors (independent variables) that statistically explain variations in some outcome criteria (dependent variables).

Some researchers who are wedded to defining process as a category of concepts may argue that one can decompose an observed sequence of events into a series of input-process-output analyses by viewing each event as a change in a variable (e.g., as the difference between nonexistence at the beginning state and existence at the ending state of the entity) and then determining whether state transitions are explained by some other independent variables. From this perspective, events represent changes in process and output variables in an input-process-output model, and the essential influence can be captured by measuring these variables and estimating the likelihood of occurrence using stochastic methods like event history analysis (Tuma and Hannan 1984). However, if the research question is *how*, not *if*, a change occurred, then an answer requires a narrative describing the sequence of events that unfolded while the change occurred. Once the sequence or pattern of events in a developmental process is found to exist, one can turn to questions about the causes or consequences of events within the process pattern.

Thus, to understand how processes of change unfold, researchers may need to alter their typical ways of modeling and methods of analysis. Rather than first generalize in terms of variables, researchers should first generalize in terms of a narrative history or a story. Only in this way will the key properties of order and sequence of events be preserved in making theoretical generalizations about processes of social change and development.

Process as a Developmental Event Sequence

A second meaning of process is a sequence of events or activities that describe how things change over time. Whereas the first definition of process examines changes in variables over time, this definition of process takes a historical developmental perspective, and focuses on the sequences of incidents, activities, or stages that unfold over the duration of an entity being studied. Table 7.2 exemplifies this meaning of process by outlining a sample of well-known process models of decision making, strategic planning, and organization development.

While the process models in Table 7.2 are concerned with the development of very different things, they are strikingly similar in two respects. First, with the exception of Cohen et al.'s (1972) garbage can model, research on all the other process models are based on cross-sectional observations or retrospective case histories in a variety of organizations. The stages or phases of activities in each model were inferred either from organizational historical self-reports or by categorizing cohorts of groups or organizations into the stages or phases. My understanding is that in no instance was any one organizational unit actually observed over time to go through all the stages or phases of any model shown in Table 7.2. Thus, there is a great need and opportunity for systematic longitudinal research to substantiate and elaborate these process models of development.

Second, in contrast with the first meaning of process as a category of variables, variables are not the centerpiece of the process models in Table 7.2.

Table 7.2. Sample of developmental process models in strategic management literature

Authors and Summaries	Beginning ←		Activity phases or stages		→	End
Strategic decision models						
Mintzberg et al. (1976) —Field study of 25 strategic, unstructured decision processes	1. Identification phase —Decision recognition routine —Diagnosis routine		2. Developmental phase —Search routine —Design routine			3. Selection phase —Screen routine —Evaluation–choice routine —Authorization routine
Cohen, March, and Olsen (1972) —Garbage can model of decision making	Decisions are probabilistic intersections of relatively independent streams within organizations of: —choices → —problems → —solutions → —energy of participants →					
Quinn (1980) —Case studies of nine major corporations	Fourteen process stages beginning with need sensing and leading to commitment and control systems. Flow is generally in sequence but may not be orderly or discrete. Some of the process stages are the following:					
	1. Sense need	2. Develop awareness & understanding	3. Develop partial solutions	4. Increase support	5. Build consensus	6. Formal commitment
Strategic planning models						
Gluck, Kaufman, and Walleck (1980) —Study of formal planning systems in 120 companies	1. Basic financial planning —meet budget		2. Forecast-based planning —predict the future	3. Externally oriented planning —think strategically		4. Strategic management —create the future
Lorange (1980) —Normative model of corporate strategic planning	1. Objectives setting —identify relevent strategic alternatives	2. Strategic programming —develop programs for achieving chosen objectives	3. Budgeting —establish detailed action program for near-term	4. Monitoring —measure progress toward achieving strategies		5. Rewards —establish incentives to motivate goal achievement
Organization development models						
Scott (1971) —Stages of corporate development	1. Single product, channel, & entrepreneurial structure		2. Single product, channel, & functional structure		3. Multiple products, channels, & division-alized structure	
Greiner (1972) —Stages of organizational growth through evolution and revolution	1. Growth through creativity —Leadership crisis	2. Growth through direction —Autonomy crisis	3. Growth through delegation —Control crisis	4. Growth through coordination —Red tape crisis		5. Growth through collaboration —Crisis of ?

Source: Ven de Ven (1992). 'Suggestions for Studying Strategy Process: A Research Note', *Strategic Management Journal*, 13: 171.

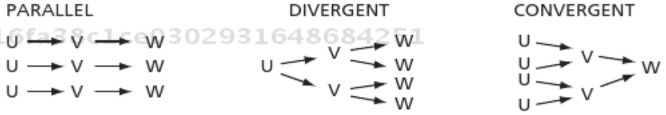
Instead, the central focus of developmental process models is on progressions (i.e., the nature, sequence, and order) of activities or events that an organizational entity undergoes as it changes over time. As the table exemplifies, a linear sequence of stages or phases of development is a common form of progression in these process models. For example, a rational process of decision making is typically viewed as a sequence of separable stages (e.g., need recognition, search, screen, and choice activities) ordered in time and with transition routines to make adjustments between stages (March and Simon 1958). Many social processes reflect far more complex progressions than simple linear sequences of stages or phases.

There are many other forms of progression that are useful for thinking about and observing developmental processes. The child development psychologists, van den Daele (1969; 1974), Riegel (1969), and Flavell (1972), for example, propose a vocabulary of developmental progressions that goes beyond simple unitary stages. As Table 7.3 illustrates, the vocabulary includes *multiple*, *cumulative*, *conjunctive*, and *recurrent progressions* of *convergent*, *parallel*, and *divergent* streams of activities as a developmental process unfolds over time. This vocabulary is useful for appreciating alternative forms of developmental progressions, which in turn, is central to understanding the

Table 7.3. A vocabulary for examining developmental progressions

Alternative Progressions of Events

- simple unitary progression
  - A sequence of the form $U \rightarrow V \rightarrow W$
- multiple progressions
  - Development can follow several paths
  - Forms: parallel, divergent, and convergent



- cumulative progressions
  - More than one stage may belong to a unit at a time.
  - Forms: by addition, substitution, or modification

$$U \supset a \rightarrow V \supset a\ b \rightarrow W \supset a\ b\ c$$
$$U \supset a \rightarrow V \supset b \rightarrow W \supset b\ c$$
$$U \supset a \rightarrow V \supset a\ b \rightarrow W \supset c$$

- conjunctive progressions
  - Events in one path are related or influence events in another path of a multiple progression
  - Relations may be probabilistic, inclusive, or mediated
- recurrent progressions
  - Repeating strings of events over time

Source: Adapted from van den Daele (1969). 'Qualitative Models in Developmental Analysis', *Developmental Psychology.*

second meaning of process. It provides the analytical terms needed to make clear distinctions between the various process models in Table 7.2.

1. A *Unitary Progression* is a simple linear sequence of the form U → V → W, where U, V, and W represent qualitatively different patterns, stages, or phases of activities or behaviors. This model assumes that each stage may consist of any number of subsets of activities, but that these subsets must occur in an ordered progression. If a developmental progression has no more than one subset of events over time, it is called a simple unitary progression, as illustrated in Table 7.2 by the two strategic planning models and Scott's (1971) stage model of corporate development.
2. *Multiple Progressions* assume that developmental processes follow more than a single path. Three common forms of multiple progressions among event sequences are the parallel, divergent, and convergent progressions illustrated in Table 7.3.

   In multiple progressions a temporal sequence of events may reflect more than one pathway at a given time in the ordered progression. For example, in the strategic decision process study of Mintzberg, Raisinghani, and Theoret (1976) in Table 7.2, more than one feasible path (or routine) of decision diagnosis, search, or evaluation might be pursued in each respective stage of identification, development, and selection. These paths diverge from each other at the beginning of each stage, proceed in parallel progressions during each stage, and converge at the end to complete each stage. As this example suggests, any developmental progression that has more than one subset of parallel paths at a time is called a multiple progression. A description of how multiple progressions of events diverge, proceed in parallel, or converge over time provides a useful vocabulary for making process statements about specific stages or the overall developmental pattern of a developing entity over time.
3. A *Cumulative Progression* (in unitary or multiple models) assumes that elements found in earlier events or stages are added and built upon in subsequent events or stages (as they are assumed to be in Lorange's (1980) and Scott's (1971) models in Table 7.2). Complete cumulation means that every event from each stage is carried from its onset until the end of the developmental progression. Of course this seldom happens, since losses of memory, mistakes and detours, and terminated pathways all imply partially cumulative or substitution progressions (as illustrated in the bottom two tracks in Table 7.3). Such partial cumulation is reflected in Quinn's (1980) 'logical incremental' model of a long sequence of 14 stages, which distinguishes it from a cumulative progression implied by a rational model of decision making.

   A cumulative progression may take the form of addition, substitution, or modification (Flavell, 1972). In *addition*, a later-occurring event supplements an earlier-occurring event. The outcomes of two events E1 and

E2 may coexist and are both equally available for E3. For example, in Scott's (1971) model of corporate development, a multiple products divisionalized structure is largely produced by the addition (with slight modification) of a stage 1 single product entrepreneurial structure with a stage 2 single product functional structure. With *substitution* the outcomes of a later event largely replace those of an earlier one. More precisely, E2 deletes or subtracts the effects of E1, and replaces them by adding those of E2. For example, in Greiner's (1972) model of organizational growth, crisis at the end of each stage leads the organization to shift (or substitute) its focus and transition into the next qualitatively new stage. In *modification* a later event represents 'a differentiation, generalization, or more stable version of the earlier one' (Flavell 1972: 345). In this case the outcome of E1 is revised or modified in E2. For example, in the strategic planning model of Gluck et al. (1980) in Table 7.2, the planning process and focus of each prior stage is modified and made more elaborate in the next stage.

4. *Conjunctive Progressions* (in unitary, multiple, or cumulative models) posit that the elements of subsets may be related. Conjunctive events are causally related events, meaning that events in one pathway may influence events in other pathways of a multiple progression. Of course what is related at one time may be viewed as unrelated at another. Therefore, strict causality among events is difficult to establish.

   Conjunctive progressions may be probabilistic, inclusive, or mediated. *Probabilistic* relationships between events occur when the trajectories of multiple paths of activities happen to intersect. Such is the form of conjunction among streams of choices, problems, solutions, and participants' energy in the garbage can model of Cohen et al. (1972). *Inclusion* occurs when the outcomes of earlier events become incorporated into the later one, as often observed with PERT charts. For example, Lorange's strategic programming phase represents the logical inclusion of alternatives from stage 1 into a strategic program in stage 2. In a *mediation* relationship an earlier event or element 'represents some sort of developmental bridge or stepping stone (mediator) to the later one' (Flavell, 1972: 345). So E2 is required in order to move from E1 to E3, which may also pre-empt alternative paths. For example, in Greiner's model crisis events mediate and bridge transitions between evolutionary stages of organizational growth.

5. *Recurrent Progressions* (in unitary, multiple, cumulative, or conjunctive models) are repeating strings of events or activities over time. Although the previous progression models have been treated as nonrecurrent sequences, parts or all of them may repeat over time. For example, what distinguishes Mintzberg's model of strategic unstructured decision processes from the others in Table 7.2 is its attention to repeating

routines, or iterative progressions, within each phase of decision making. Abbott (1990) discusses a variety of techniques for the colligation and measurement of recurrent and nonrecurrent event sequence data.

These alternative models of progression in Table 7.3 do not occur independently. Whether implicit or explicit, every development process model makes a commitment to some form of temporal progression of unitary or multiple sequences of events that may be cumulative, conjunctive, and reoccur over time. This vocabulary of temporal relationships among events can help scholars articulate the meanings of their process models in more operational and discriminating ways than in the past. However, this analysis of process as a sequence of events cannot go far without considering the alternative theories of process that may explain specific developmental progressions.

CLARIFY THEORIES OF PROCESS

Whereas a definition of process indicates one's meaning of process in relation to other uses, a theory of process consists of an explanation of how and why a process unfolds over time. Such a theory is useful not only to ground the conceptual basis of a process study, but also to guide the design and conduct of an empirical study. Thus, the second basic decision for designing a process study is to clarify the theory of process underlying the substantive investigation.

I do not wish to imply that you have a clear process theory in mind before undertaking empirical research so that it can be tested. In my experience, I have never been sure what process theory might be useful to explain field observations. It is precisely because of this ambiguity in not knowing what to expect that a repertoire of alternative models is immensely helpful in making sense of reality. As Pasteur advised, 'Chance favors the prepared mind.'

Viewing process as a developmental progression, Scott Poole and I proposed four basic theories that serve as ideal types for explaining processes of development and change in organizations (Van de Ven and Poole 1995). Figure 7.1 shows that each theory views the process of development as unfolding in a fundamentally different progression of change events, and to be governed by a different generative mechanism or motor.

- A *life cycle* (or regulated) model depicts the process of change in an entity as progressing through a necessary sequence of stages or phases. In terms of the vocabulary introduced before, the typical progression of a life cycle process of change is a unitary, cumulative, and conjunctive sequence of stages, because the content and historical sequence of these stages is

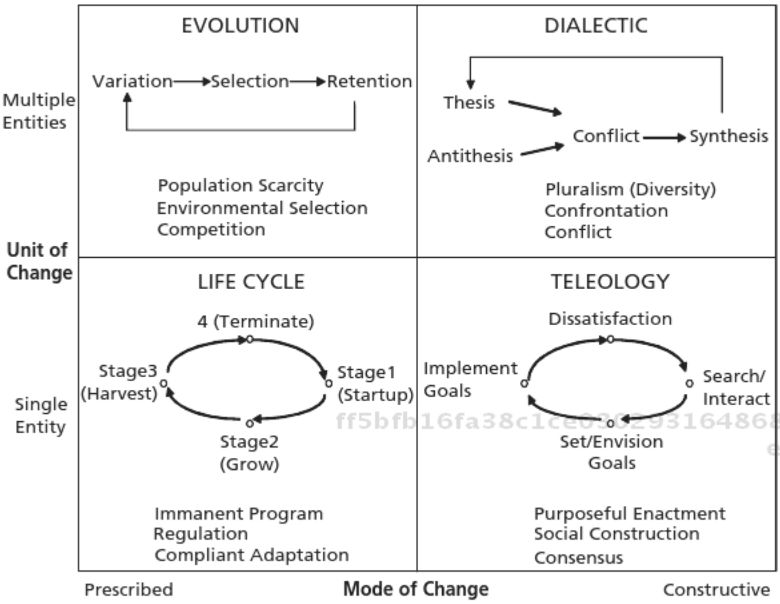


Figure 7.1. Process theories of organizational development and change

Source: Van de Ven, A. H. and Poole, M. S. (1995). 'Explaining Development and Change in Organizations,' *Academy of Management Review*, 20(3): 520.

Note: Arrows on lines represent likely sequences among events, not causation between events.

prescribed and regulated by an institutional, natural, or logical program prefigured at the beginning of the cycle.

- A *teleological* (or planned change) model views development as a cycle of goal formulation, implementation, evaluation, and modification of actions or goals based on what was learned or intended by the entity. This sequence emerges through the purposeful enactment or social construction of an envisioned end state among individuals within the entity. Teleological models of development incorporate the systems theory assumption of equifinality; there are several equally effective ways to achieve a given goal. There is no assumption about historical necessity. Rather, these models rely on agency as the explanatory principle: they posit a set of functions or goals desired by an organizational unit, which it has to acquire in order to 'realize' its aspirations. Development is movement toward attaining a purpose, goal, function, or desired end state.
- In *dialectical* models of development conflicts emerge between entities espousing an opposing thesis and antithesis that collide to produce a synthesis, which in time becomes the thesis for the next cycle of a dialectical progression. Confrontation and conflict between opposing agents generate this dialectical cycle. Stability and change in a dialectical process theory are

explained by the relative balance of power between opposing forces. Stability is produced through partisan struggles and accommodations, which maintain the status quo between oppositions. Change occurs when these opposing values, forces, or events go out of balance. The relative strength, power, or legitimacy of an antithesis may emerge or mobilize to a sufficient degree of force to overthrow the current thesis or state of affairs and produce a synthesis, which then becomes the new thesis as the dialectical process recycles and continues.

- An *evolutionary* model explains change as a recurrent, cumulative, and probabilistic progression of variation, selection, and retention among entities in a designated population. This evolutionary cycle is generated by competition for scarce environmental resources between entities inhabiting a population. As in biological evolution, change proceeds in a continuous process of variation, selection, and retention. Variations, the creation of novel forms, are often viewed to emerge by blind or random chance; they just happen. Selection occurs principally through the competition among forms, and the environment selects those forms that optimize or are best suited to the resource base of an environmental niche. Retention involves the forces (including inertia and persistence) that perpetuate and maintain certain organizational forms. Retention serves to counteract the self-reinforcing loop between variations and selection.

Two dimensions are useful for distinguishing the four process models illustrated in Figure 7.1: (1) whether the unit of change involves one or more entities; and (2) whether the mode of change is prescribed or constructed. Life cycle and teleological theories operate on a *single entity*. In the case of a life cycle model, the development of any entity is governed by a code immanent within the entity or a set of institutional rules to which the entity adapts while changing. While the environment and other entities may shape how an entity adapts, they are strictly secondary to the immanent forces for development within the single entity. Teleological theories also focus on only a single entity's goals, social construction, or envisioned end state to explain development. A teleological theory can operate among many members of an organization or a set of organizations when there is sufficient consensus among the members to permit them to act as a single organizational entity. On the other hand, evolutionary and dialectical theories operate on *multiple entities*. Evolutionary forces are defined in terms of their impact on populations and have no meaning at the level of the individual entity. Dialectical theories require at least two entities to fill the roles of thesis and antithesis.

The generative mechanisms of the four process theories also differ in terms of a second dimension regarding whether the sequence of change events is prescribed *a priori* or whether the progression is constructed and emerges as the change process unfolds. A *prescribed* mode of change channels the development of entities in a pre-specified direction, typically of maintaining and incrementally adapting their forms in a definite, calculable way. A *constructive* mode of change generates unprecedented, novel forms that, in retrospect, are often discontinuous and unpredictable departures from the past. A prescribed motor evokes a sequence of change events in accord with a pre-established program or action routine. A constructive motor, on the other hand, produces new action routines that may (or may not) create an original (re)formulation of the entity. Life cycle and evolutionary theories operate in a prescribed modality, while teleological and dialectical theories operate in the constructive modality.

Most researchers conduct their studies with one model or theory in mind. Working with a single model or perspective of change has the advantage of sharpening and focusing data collection and analysis. A single perspective or model is also easier to operationalize and fit the data. However in Chapter 4, I argued, in contrast, that having two or more models enables the researcher to make stronger inferences by positing a series of critical tests of assumptions that differentiate the models. Another advantage of comparing plausible alternative models is that null results on one model are less likely to leave the researcher in a cul-de-sac of knowing only what is not the case.

Most organizational change processes can be exceedingly complex, and far beyond the explanatory capabilities of any single process theory found in the literature. Typically several different models are needed to capture different aspects of the same process; they complement each other to better understand the process (Pettigrew 1990). Moreover, when researchers and practitioners have only a single perspective or theory, they tend to twist and rationalize facts to fit their model (Mitroff and Emshoff 1979). Consequently, I suggest it is generally better to develop and juxtapose alternative theories and then determine which theory better explains the data or how they can be combined.

The comparative method also facilitates keeping the research focused and manageable. It reduces complexity because it is very difficult to analyze a large array of field data without conceptual guidance. This approach emphasizes that testing a process theory should be based on the relative explanatory power of alternative theories that are available or that can be developed to explain the phenomena. It is also consistent with the principle that knowledge advances by successive approximations and comparisons of competing alternative theories (Lakatos 1978).

FRAME OF REFERENCE TO VIEW THE RESEARCH QUESTION

Once the meanings and theories of process are clear, then a researcher has the basic conceptual foundations for designing a process study undertaken to examine a specific research question about how change unfolds over time.

A crucial step in launching any study is being reflexive about the researcher's role and perspective. As discussed in Chapter 2, a researcher can only observe and recount a partial view of the events that may unfold in a change process (Schein 1987). The view that scientific observations can be impartial or detached has been severely discredited (Popper 1972). Most social scientists now concede that no research is value-free; a researcher should therefore disclose his/her values and perspective (Van Maanen 1995; Alvesson and Skoldberg 2000).

Every act of observing something represents countless choices not to observe other things and perspectives. Any topic or issue can be examined from the viewpoints of many different individuals or stakeholders. Some of these viewpoints are accessible to the researcher, others are not. It is difficult, if not impossible, for a researcher to assume an impartial and detached perspective or to obtain a balanced representation of all stakeholders involved in any complex organizational change process. It is better to be explicit about which stakeholder's interests and viewpoints are favored (and accessible) than to be silent or naïve about whose interests are served and ignored in any study.

Following this recommendation, engaged scholars often aim to see organizational life from the perspective of a specific participant or stakeholder in the process. This often requires more than a detached view of the subject; indeed, researchers may actively participate in the lives of the people and situations that they are studying (Singleton et al. 1993).

This requires a degree of access and engagement with key stakeholders that few researchers have been able to develop. Gaining access is problematic for many researchers because they seldom place themselves into the frame of reference of the stakeholders who sponsor the study or wish to use its results. Typically, managers are key stakeholders in field studies of change in their organizations. Without observing a change process from the manager's perspective, it becomes difficult for a researcher to understand the dynamics confronting managers who are directing the change effort, and thereby generate new knowledge that advances the theory and practice of managing change. If organizational participants do not understand the relevance of a study, there is also little to motivate them to provide access and information to an investigator. The issue here is *not* that researchers become consultants. As discussed further in Chapter 9, the issue is one of engaging key participants in a study in formulating important research questions that capture the attention and motivation of scholars and practitioners alike.

For example, in launching the Minnesota Innovation Research Program (MIRP) (Van de Ven et al. 2000), we found that a useful way to begin formulating a longitudinal field study was to conduct periodic meetings with small groups of managers from various organizations engaged in comparable change efforts or new ventures. In these meetings we discussed the meanings and implications of the research question (e.g., How and why do innovations develop over time?) and explored ways of studying the question so that it might advance theory and practice from a manager's viewpoint. These meetings produced many useful ideas that guided our research, and many participants also agreed to provide access to conduct the research. Moreover, these meetings often identified individuals whom we negotiated with to become study advisors, facilitators, or co-investigators.

MODE OF INQUIRY

Reflecting on their styles of inquiry and clarity of the subject matter, researchers can adopt a continuum of strategies that are grounded in theory or data. While *deduction,* a theory-driven approach, is familiar to most readers, *abduction*, and its relationship to the more popular term, *induction*, may not be. As discussed in Chapter 4, induction refers to the inference we draw from direct observation of a phenomenon that results in assigning a probability of the likelihood of an occurrence in the future. Abduction refers to a conjecture or hypothesis that we invent to explain anomalies or surprising patterns that we observe (Peirce 1955). Such a conjecture or hypothesis should go beyond the information given in a specific case (Bruner 1973). Since abduction more accurately describes the mode of reasoning entailed in grounded theorizing than induction, I use the term abduction instead of induction.

With a deductive approach, the basic steps in designing research might consist of adopting one or more process theories of change (e.g., Figure 7.1), developing an operational template for the theory, and then using it to determine how closely an observed process matches the theory. With abduction, the steps might include observing processes of stability and change over time in a few organizational entities, sorting data into meaningful categories, developing propositions explaining the observations, and corroborating them with a different sample or on the same sample at a different time.

There is a tight iterative cycle between deduction, abduction, and verification in grounded theory building studies. Strauss (1987) emphasized that all scientific theories require that they be conceived, then elaborated, then checked. 'Few working scientists would make the mistake of believing these stood in a simple sequential relationship.... Many people mistakenly refer to grounded theory as "inductive theory"... All three aspects of inquiry (induction, deduction, and verification) are absolutely essential' (Strauss 1987: 11–12). In the course of a longitudinal study, most researchers move back and forth between these modes of inquiry many times.

OBSERVING PROCESSES IN REAL TIME OR RELYING ON RETROSPECTIVE ACCOUNTS

Because change is defined as an observed difference in an organizational entity over time, a process study necessarily entails collecting longitudinal data. These data can be obtained either by observing the sequence of change events as they occur in real time, or by relying on archival data to obtain a retrospective account of the change process. Most studies of organizational change are retrospective, conducted after outcomes are already known before data collection begins. Retrospective studies provide the advantage of knowing the 'big picture'—how things developed and the outcomes that ensued. This *post hoc* knowledge is helpful for interpreting events that unfolded, and for constructing a narrative of the process. When researchers conduct real-time observations of a change process as it unfolds, they do not have this advantage of afterthought and may miss occurrences or events that later can be viewed as critical. Until we have the compass of the entire process, we often have no way of knowing what information is important and what is not.

However, prior knowledge of the outcome of an organizational change may also bias a study. This is especially true if the final assessment valorizes the outcome as a success or failure, effective or ineffective. There is a tendency to filter out events that do not fit or that render the story less coherent, such as censoring minority views.

A promising approach is to initiate historical study before the ultimate outcomes of a change process become apparent. It is even better to observe the change process in real time as it unfolds in the field setting. This approach maximizes the probability of discovering short-lived factors and changes that exert an important influence. As Pettigrew (1985) notes, 'the more we look at present-day events, the easier it is to identify change; the longer we stay with an emergent process and the further back we go to disentangle its origins, the more likely we are to identify continuities.' At one point or another, most field studies of organizational change involve many forms of longitudinal data collection: archival, retrospective, and real-time observations.

SOURCES OF CHANGE

In the study of human development, Schaie (1965) discussed three common sources of temporal change:

1. *Age*: The age or temporal duration of the individual at the time of measurement. This variable represents that part of development and change that is produced by unfolding biological or institutional processes.

2. *Cohort*: The set of characteristics of all individuals who were born at the same time and go through similar developmental processes, such as classes in school. This variable represents the common historical conditions that shape the development of a given cohort.
3. *Transient*: All the temporary or immediate and non-cumulative factors that influence outcomes or the dependent variables at the time of measurement.

Schaie suggests that it is important to design organizational change studies so they can disentangle these three sources of change—those that are due to age, to external factors in the history of the developing organism (cohort), or to immediate external factors (time of measurement). What appears to be a developmental change due to some immanent mechanism could well be due to a cohort effect or to a unique effect at the time of measurement. For example, a sudden shift in morale compared to previous levels may result from a general improvement in social mood at the time of measurement. Interpreting this as a function of solidification of a developing culture would be incorrect, though it would be easy to see why a researcher whose attention focused only on the organization under study might draw this conclusion. In the same vein, what appears to be a general developmental pattern might be due to cohort effects, unique events occurring only to the group of organizations that were founded in a given time and place. By this reasoning, for example, it would be risky to try to generalize principles of effective development of organizational start-ups in the relatively benign 1950s to organizations in the competitive 1990s because they belong to different cohorts. They operated and started under different resource constraints, had employees with different attitudes, and had a different external environment.

This is not to imply that it is impossible to develop generalizable findings concerning development and change. Rather, it is important to consider what source observed changes may originate from and to rule out alternative explanations for the ones we advance. It is also important to consider the limits of our conclusions. Taking into account age, cohort, and time of measurement as well as organization type and context will result in more effective research designs.

Barley's (1990) research design, shown in Figure 7.2, provides a good example of a systematic study of these different sources of change. In his field study of the adoption of a technology (CT scanners), Barley drew comparisons between two parallel hospitals with synchronic (one point in time) observations of different radiology technologies, and with diachronic (repeated over time) observations of CT scanning behavior by radiology department staff. Reflecting on his design, Barley discusses how conclusions can become problematic when the research questions and comparative analysis are not matched correctly.

For example, synchronic data may seem to suggest that similar outcomes are rooted in similar processes. However, similar outcomes may arise from different processes and different outcomes may arise from similar dynamics (Barley, 1990: 186). Only diachronic data can disentangle such possibilities. By itself, a parallel study of a class of events, objects, or activities may also lead to wrongful conclusions. Suppose, for instance, that one were to investigate the effects of new technologies by studying CT scanning in a number of hospitals. Even if one found that all CT scanners occasion similar phenomena, one could not be sure whether the findings would apply to all computationally based imaging devices or only to CT scanners. A synchronic analysis of several technologies conducted in tandem could resolve this issue. In other words, the synchronic, the diachronic, and the parallel represent three distinct axes of comparison that, when used in combination, allow researchers to examine explicitly the spatial and temporal boundaries of their claims. (Barley 1990: 227)

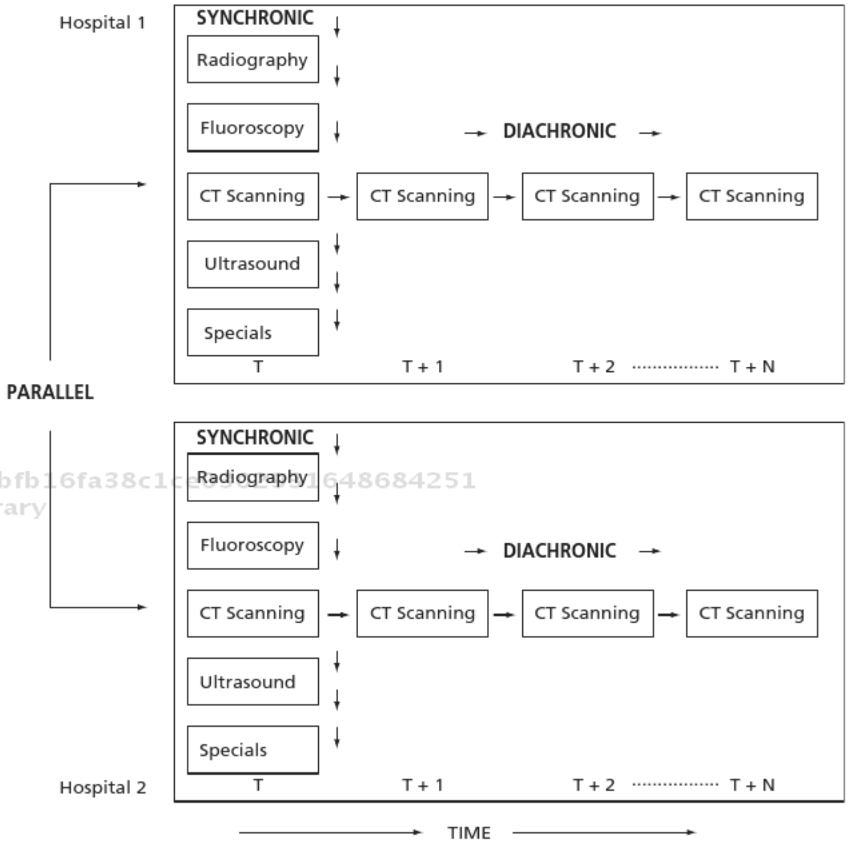


Figure 7.2. Barley's (1990) parallel, synchronic, and diachronic research design

Source: Barley, S. R. (1990). 'Images of Imaging: Notes on Doing Longitudinal Research,' *Organization Science*, 1(3): 226.

SAMPLE DIVERSITY: HOMOGENEOUS OR HETEROGENEOUS CASES

There is no one best sampling scheme for process research. A homogeneous sample has the advantage of keeping to a minimum the multitude of alternative explanations for developmental processes. This is especially advantageous in the case of lengthy sequences of events, because they are particularly vulnerable to accidental or adventitious occurrences that shift the course of development. Comparing cases that are similar in as many respects as possible facilitates identifying whether change processes are due to such transient events or to more basic developmental models, but does not control for cohort effects. A homogeneous sample also facilitates the development and investigation of very precise, focused questions or hypotheses. Hence homogeneous sampling is useful when a well-specified theory of change or development is available. A broad, heterogeneous sample, however, may provide a better opportunity to detect whether sources of change are due to temporal development, cohort, or transient factors.

The comparative method is perhaps the most general and basic strategy for generating and evaluating valid scientific knowledge. This strategy involves the selection of comparison groups that differ in the scope of the population and conceptual categories of central interest to the research. Kaplan (1964: 52) pointed out that scientific knowledge is greatly enhanced when we divide the subject matter into concepts and cases that 'carve at the joints' over the widest possible ranges, types, conditions, and consequences. In this way researchers can develop and evaluate the limits of their propositions.

A broad sampling scheme also permits a researcher to make empirical links between different specialties or schools of thought that have emerged for different organizational settings in which the change process occurs. For example, because organizational structures for business creation are different in small company start-ups, internal corporate innovation projects, and inter-organizational joint ventures, it is widely believed that the process of entrepreneurship in these organizational settings must also be different. Our MIRP studies questioned this conventional belief, and proposed the plausible alternative that creating a new business entails fundamentally the same process regardless of organizational setting. We obtained some empirical evidence supporting this proposition (Van de Ven et al. 1999). The findings suggest that significant benefits and efficiencies can be gained by applying principles of business creation from new company start-ups to internal corporate venturing and inter-organizational joint ventures, and vice versa.

Given the tradeoffs between homogeneous and heterogeneous samples, Pettigrew (1990: 275–7) suggests four useful guidelines for selecting cases to study:

1. 'Go for extreme situations, critical incidents and social dramas.' By choosing unusual cases, cases that are critically important or highly visible cases, researchers select cases in which the process is 'transparently observable.' However, such cases may have nongeneralizable features precisely because they are unusual.
2. 'Go for polar types.' Choose cases that seem very different in terms of the processes under study. For example, compare successful and unsuccessful program start-ups. Or, choose cases that differ from patterns in earlier cases. By successive sampling of polar types, it will eventually be possible to cover the range of possible cases.
3. 'Go for high experience levels of the phenomena under study.' Choose cases that have a long track record of experience with a process. This strategy may not be feasible for some cases: new program start-ups, for example, may best be illuminated by inexperienced entrepreneurs, since they will make the mistakes and experience the learning that highlights key requirements for successful start-ups.
4. 'Go for a more informed choice of sites and increase the probabilities of negotiating access.' Cases must often be selected on the basis of who will cooperate, rather than on grounds of optimal sampling. This, of course, introduces a sampling bias that must be considered in drawing conclusions from the study.

SAMPLE SIZE: NUMBER OF EVENTS AND/OR CASES

The major sample size consideration in variance research studies is the number of cases selected for data collection, as discussed in Chapter 6. The larger the number of cases that are sampled from a population of interest, the more generalizable are the results (provided that the cases are drawn in a representative fashion). Furthermore, in experimental designs, researchers are advised to select the number of cases needed to obtain enough power for statistical tests to equate statistical significance with practical significance in hypotheses testing (Walster and Cleary 1970). Pragmatically, the number of cases selected also depends on the availability of sites and the costs involved in collecting data on each case.

In longitudinal process studies, the central sample size consideration is the *number of temporal intervals or events* obtained on a change process in each case. The number of temporal intervals or events observed depends on what constitutes the 'natural' flow of experience in the organizational change cases being studied. Organizational change processes vary in *temporal duration* and *granularity*. In terms of temporal duration, some organizational change processes, such as group decision making, may occur in committee meetings lasting no more than a few hours. Other change processes, such as the development of technological and administrative innovations, may span several years.

Granularity refers to the preciseness or discreteness of events that are recorded throughout the temporal duration of a case being studied. The granularity of events varies greatly, ranging from events of such large scope that only 5 to 20 might be observed over the period of study to events of such small scope that several thousand occur. Event granularity typically increases with the micro-analytic detail of the change process being investigated.

Events that require a great amount of time and effort to observe and code are likely to be observed in shorter sequences than those less costly to observe. Because there are inherent tradeoffs between the temporal duration and granularity of events that can be sampled, most studies of fine-grained events tend to focus on change processes of relatively short temporal duration, while studies of lengthy change processes tend to adopt coarse-grained events.

PROCESS RESEARCH DESIGNS

There are important implications of the number of cases and events observed in a study for process research design and data analysis. Poole et al. (2000) discuss these implications with reference to their typology of alternative process research designs shown in Table 7.4.

Studies consisting of *few cases, few events* reflect the typical sampling design of comparative case studies. Sometimes there may be few events, not due to paucity of data, but because only a few occur. For example, in a comparative study of strategic decision making where the sequence of search, screen, and choice behaviors are being investigated, there may be relatively few instances of each type of behavior in the case. Alternatively there may be only a couple of instances of the key events (e.g., conflicts) in otherwise lengthy cases. Provided there are enough cases for systematic comparison and induction across the instances, Yin's (2003) comparative case study designs can be utilized.

Table 7.4. Typology of process research designs from Poole et al. (2000)

	FEW EVENTS	MANY EVENTS
Few cases	Summary case studies	Summary case studies Phasic case studies Time series analysis Markov analysis
Many cases	Multivariate analysis Phasic analysis with optimal matching Event history analysis	Multivariate analysis of summary data Phasic analysis with optimal matching Markov analysis Time series analysis

Studies with *many cases, few events* provide many comparative options for the researcher. *Summary measures* for each case can be derived by collapsing the data along the time dimension (e.g., counting the number of conflicts that occur during innovation regardless of when they occurred), or through the use of surrogate measures of temporal order (e.g., did the conflict occur during the first or second halves of the innovation process?). Such measures can then be treated as variables in traditional statistical methods. However, with such pooling of the data, one can lose the temporal order of events that figure prominently in most process research studies.

One method to preserve information about temporal order that clusters cases with similar sequences is *optimal matching*. Poole et al. (2000) discuss that once clusters have been derived, they can serve as the basis for variables that can then be entered into traditional statistical analyses. Alternatively, Tuma and Hannan (1984) discuss how *event history* or *survival analysis* can be used to determine when critical events occur, provided the length of time until they occur is recorded. Supplementary analysis can in some cases divulge causal factors underlying event occurrences (Willett and Singer 1991).

A different set of options are open for studies with *few cases, many events*. Comparative analysis of *qualitative case studies* using Yin's designs are one option. Events can be parsed into *phases* representing coherent periods of activities subsuming two or more events in sequence. These phases can then be used as bounded units to provide temporal divisions in case studies, as Holmes (1997) did in his studies of hostage-taking situations and Van de Ven and Polley (1992) did in their study of a biomedical innovation. Various types of *time series* analyses can also be used when many events are available for each case. These generally involve transforming the event series into some continuous form. In addition, Poole et al. (2000) discuss the application of *Markov analysis*, which preserves the categorical qualities of the event series and enables us to track temporal dependencies among events.

For studies with *many cases, many events* a number of powerful statistical techniques are available. As with the many cases, few events situation, simple *descriptive summaries* of the frequency with which coded events occur provide useful displays for examining stages or phases in the developmental progression. However, with such pooling of the data, one can lose the temporal order of events that figure prominently in most process research studies. *Optimal matching* can be used to derive measures of similarity among the event sequences for the cases and these measures can then be analyzed in at least two ways. First, they can be used as input to cluster analysis and multidimensional scaling techniques that can identify clusters of similar sequences; the resulting clusters can then be used to define variables for causal or correlational analysis, as in Poole and Holmes (1995). Second, these distances can be used to test for causal factors that create the differences between pairs of sequences. Poole et al. (2000) also discuss how *trend analysis* or *multiple* *time series methods* can be used to identify patterns of change across many cases, provided the events can be used to define continuous variables. *Markov analysis* of multiple cases can provide maps of temporal dependencies among events. Causal factors leading to such dependencies can then be analyzed using *Markovian regression* techniques or other simpler designs.

Measuring and Analyzing Process Data

At the heart of any longitudinal study is measuring and analyzing process data. This section reviews techniques for gathering, tabulating, and analyzing process data. In a typical longitudinal field study, the gathering of data might entail the following procedures:

- survey questionnaires completed by all participants every six months;
- interviews with key managers and participants every six months;
- direct observations of regularly scheduled meetings;
- a diary recording informal discussions with participants; and
- documents and reports from news media and organizational archives.

Whatever data collection methods are used to observe change processes in the field or from archival records, over time data mount astronomically and overload the information processing capacity of even the most insightful mind. Drawing careful inferences requires methods that go beyond subjective 'eyeballing' of raw data to identify patterns. But it is difficult to reconstruct field methods, because they are rarely reported in detail in published field studies. One cannot ordinarily follow how the researchers arrived at their conclusions from hundreds of pages of field observations, even though the reports may be sprinkled with vivid—yet idiosyncratic—quotes from organizational participants. As in variance research, methods for measuring and analyzing process data require explicit and careful attention. Chapter 6 discussed well-established psychometric procedures for survey instrument construction and evaluation. The remainder of this chapter deals with analogous, but less well-established procedures for measuring and evaluating process data. These procedures and decisions are outlined at the bottom of Table 7.1.

PROCESS CONCEPTS

Whether a researcher sets out to develop or test a process theory, the collection of longitudinal data requires a set of categories or concepts. These concepts provide selective focus for observing a change process;

one cannot study everything, and different categories can produce very different findings. When a particular process model(s) is proposed or known beforehand, category development proceeds deductively by operationalizing theoretical constructs into empirical indicators of those constructs. When a grounded theory building approach is taken, these initial categories are best viewed as 'sensitizing constructs' for conducting exploratory research. The categories become clear as they are grounded in field observations. Eventually, these grounded concepts can be codified in a final category system.

A grounded theory-building strategy provides a useful first step in developing some basic concepts and ideas from raw data. To its originators, Glaser and Strauss (1967) and Strauss and Corbin (1990), grounded theory building consists of the following structured steps. Begin with small units of data (incidents) and gradually construct a system of categories or concepts that describe the phenomena being observed. The categories may have several subcategories and dimensions that are gradually elaborated and refined as specific incidents are examined, coded, and compared. As the categories are developed, additional data are examined to verify the properties of the emerging category system. The analysis concludes with the identification of a small number of core categories that serve to integrate the theoretical concepts that are firmly rooted or 'grounded' in the data.

In our Minnesota Innovation Research Program (MIRP), for example, we began with five 'sensitizing categories' to study innovation development: ideas, people, transactions, context, and outcomes (Van de Ven et al. 2000). As is typical in longitudinal studies, our assumptions and definitions of these

Table 7.5. Evolution of innovation concepts during MIRP

	Starting definitions from literature	But we see this in field studies
Ideas	One invention to be operationalized	Reinvention, proliferation, reimplentation, discarding, and termination of many ideas
People	An entrepreneur with a fixed set of full time people over time	Many entrepreneurs, distracted, fluidly engaging, and disengaging in a variety of roles over time
Transactions	Fixed network of people/firms working out the details of an innovative idea	Expanding and contracting network of partisan stakeholders converging and diverging on innovation ideas
Context	Environment provides opportunities and constraints on innovation process	Innovation process constrained and created by multiple enacted environments
Outcomes	Final result orientation: a stable order comes into being	Final results may be indeterminate; multiple in-process assessments and spinoffs; Integration of new order with the old
Process	Simple cumulative sequence of stages and phases of development	From simple to multiple progressions of divergent, parallel, and convergent paths; some are related and cumulative, others not

Source: Van de Ven et al. (1999).

concepts over time changed substantially and became progressively clear with field observations. Table 7.5 compares our starting assumptions of these concepts drawn from the literature at the time, with how we came to view them as a result of two years of field studies. The latter disclosed a different reality from the rather orderly and naïve conceptions of the former. As this example illustrates, the development of research constructs involves an iterative process of conceptualization, observation, and reformulation.

INCIDENTS AND EVENTS

It is useful to distinguish between *incidents* and *events* in a process theory (Abbott, 1984), which are analogous to the distinction between *variables* and *constructs*, respectively, in a variance theory (discussed in Chapter 6). Incidents are operational empirical observations, while events are abstract concepts of bracketed or coded sets of incidents. The stream of incidents, a directly observable first-order set of activities, is translated into a sequence of events, a more abstract second-order construction. This implies that some incidents may be embedded in different conceptual domains and utilized as constituents of different events.

Events may differ in temporal and spatial scope, and as a result, incidents may indicate more than one, overlapping event. For example, a meeting with 'firm Q' can indicate the event 'meeting with a partner,' but it may also indicate a longer event, 'negotiation with firm Q regarding partnership.' Events may be embedded within different types of events of a larger scope. Both levels may be important for understanding the change process, because interwoven narratives clarify it better than either narrative could on its own. Abbott (1992) gives an example from his studies of the rise of professions in society, 'I once set out to explain why there are no psychiatrists in American mental hospitals. The exodus, which dates from 1900–30, reflects not only the rational individual mobility decisions that are specifiable annually, but also outpatient community developments that are specifiable only decadely, and changes in knowledge and social control taking place over even longer periods.'

Another complication is the possibility that the incident–event relationship may change over time (Abbott 1984). The significance of events may change as the process unfolds. The same change is possible in incident–event relations. For example, the first time a potential partner is encountered may signal an expansion of an organizational program, whereas the sixth encounter with a potential partner may signal desperation for ideas or resources. Thus, while events are constructs indicated by incidents, the indication relationship is more complicated for qualitative data than it is for quantitative scores. The assumption of uniformity across respondents and responses in psychometrics and scale theory may not hold for data used to define events. What quantitative analysis would classify as an error may be quite important nuances for qualitative data.

DEFINING AN INCIDENT: A QUALITATIVE DATUM

In survey research, a *quantitative datum* is commonly regarded to be: (1) a numerical response to a question scaled along a distribution; (2) about an object (the unit of analysis); (3) at the time of measurement; which is (4) entered as a variable (along with other variables on the object) into a record (or case) of a quantitative data file; and (5) is subsequently recoded and classified as an indicator of a theoretical construct.

In comparison, we define a *qualitative datum* as: (1) a bracketed string of words capturing the basic elements of information; (2) about a discrete incident or occurrence (the unit of analysis); (3) that happened on a specific date; which is, (4) entered as a unique record (or case) in a qualitative data file; and (5) is subsequently coded and classified as an indicator of a theoretical event.

The basic element of information in a qualitative datum is a bracketed string of words about a discrete incident. Raw words, sentences, or stories about incidents that are collected from the field or from archives cannot be entered into a qualitative data file until they are bracketed into a datum(s). Obviously, explicit decision rules that reflect the substantive purposes of the research are needed to bracket raw words.

In our MIRP studies, the decision rule used to bracket words into a qualitative datum was the definition of an incident that occurred in the development of an innovation (Van de Ven et al. 2000). An incident occurred whenever changes were observed to occur in any one of our five core concepts: innovation ideas, people, transactions, context, and outcomes. When an incident was identified, the bracketed string of words required to describe it included: date of occurrence, the actor(s) or object(s) involved, the action or behavior that occurred, the consequence (if any) of the action, and the source of the information. As with any set of decision rules, discussions among researchers were necessary to define innovation incidents in an operationally consistent manner.

Decision rules may vary in the level of specificity and the temporal duration of incidents they construct. Some rules specify fine-grained definitions of incidents that interpret each action as a separate incident; others adopt coarse-grained definitions that require longer episodes for incidents. The proper granularity of incidents depends on the rates of development of various kinds of processes, and the differing research questions associated with these rates.

For example, Knudson and Ruttan (2000) found that hybrid wheat development was governed by biological laws that require several decades to move from basic research through technology development to market introduction. They observed that hybrid wheat's innovation process had been following this 'biological time clock' for forty years since the late 1950s. In studies of biomedical innovations, Garud and Van de Ven (2000) observed that the rate of development was governed by an 'institutional regulation time clock,' in which the design, testing, and commercial release of devices entailed extensive review and approval steps by the US Food and Drug Administration, sometimes lasting five years. However, rates of development of other processes, such as group decision making (Poole and Roth 1989) or the development of novel administrative programs (Roberts and King 1996; Bryson and Roering 2000) are more rapid and appear to be limited only by entrepreneurial time and attention. As these variations suggest, the temporal scope of organizational change should correspond with the granularity of incidents being observed in the field study. Zaheer et al. (1999) provide a stimulating discussion of these and other considerations in developing temporal metrics.

RELIABILITY AND VALIDITY OF INCIDENT CONSTRUCTION

It is important to establish the reliability of classifying raw data into incidents. An equally important, though often neglected, issue is the validity of this bracketing procedure (Folger et al. 1984; Poole et al. 1987). Researchers often assume that the meaning of incidents is clear, and that establishing reliability is equivalent to showing the meaning of codings is clear. However, attaining reliability among coders simply indicates that the meaning of incidents is clear to the particular group of researchers who designed the coding system, not necessarily to participants or key stakeholders. It is necessary to test empirically whether researchers' classifications coincide with practitioners' perceptions of events. If the evidence indicates inconsistency, then no claims about the meaning of events to the participants are valid. Researchers can still sustain claims about the meaning of the incident from their theoretical position, but no claims about the 'social reality' of the event are appropriate.

Two basic procedures can enhance the reliability and validity of incident coding. First, coding of incidents from raw data sources can be performed by two or more researchers. Consensus among coders increases the consistency of interpretations of the decision rules used to identify incidents. Second, incident codings can be reviewed by key organizational informants. It is useful to ask informants if any incidents are missing or incorrectly described. Based on this feedback, revisions in the incident listings can be made if they conform to the decision rules for defining each incident. Typically, these two steps