

Chapter 4

Phase One: Designing Case Study Research

There are three phases in the design and implementation of theory-oriented case studies. In phase one, the objectives, design, and structure of the research are formulated. In phase two, each case study is carried out in accordance with the design. In phase three, the researcher draws upon the findings of the case studies and assesses their contribution to achieve the research objective of the study. These three phases are interdependent, and some iteration is often necessary to ensure that each phase is consistent and integrated with the other phases.¹⁵⁷ The first phase is discussed in this chapter, and phases two and three in the chapters that follow.

Phase one—the research design—consists of five tasks. These tasks are relevant not only for case study methodology but for all types of systematic, theory-oriented research. They must be adapted, of course, to different types of investigation and to whether theory testing or theory development is the focus of the study. The design phase of theory-oriented case study research is of critical importance. If a research design proves inadequate, it will be difficult to achieve the research objectives of the study. (Of course, the quality of the study depends also on how well phases two and three are conducted.)

Task One: Specification of the Problem and Research Objective

The formulation of the research objective is the most important decision in designing research. It constrains and guides decisions that will be made regarding the other four tasks.

The selection of one or more objectives for research is closely coupled with identification of an important research problem or “puzzle.” A clear, well-reasoned statement of the research problem will generate and focus the investigation. A statement that merely asserts that “the problem is important” is inadequate. The problem should be embedded in a well-informed assessment that identifies gaps in the current state of knowledge, acknowledges contradictory theories, and notes inadequacies in the evidence for existing theories. In brief, the investigator needs to make the case that the proposed research will make a significant contribution to the field.

The research objective must be adapted to the needs of the research program at its current stage of development. Is there a need for testing a well-established theory or competing theories? Is it important to identify the limits of a theory’s scope? Does the state of research on the phenomenon require incorporation of new variables, new subtypes, or work on different levels of analysis? Is it considered desirable at the present stage of theory development to move up or down the ladder of generality?¹⁵⁸ For example, as noted in Chapter 2, in the 1990s the democratic peace research program moved largely from the question of whether such a peace existed to that of identifying the basis on which democratic peace rests. It now needs to go further to explain how a particular peace between two democratic states developed over time. Similarly, in the 1960s deterrence theory needed to bring in additional variables to add to excessively parsimonious and abstract deductive models.

In general, there are six different kinds of theory-building research objectives. Arend Lijphart and Harry Eckstein identified five types. We outline these below and add a sixth type of our own:¹⁵⁹

- *Atheoretical/configurative idiographic* case studies provide good descriptions that might be used in subsequent studies for theory building, but by themselves, such cases do not cumulate or contribute directly to theory.
- *Disciplined configurative* case studies use established theories to explain a case. The emphasis may be on explaining a historically important case, or a study may use a case to exemplify a theory for pedagogical purposes. A disciplined configurative case can contribute to theory testing because it can “impugn established theories if the theories ought to fit it but do not,” and it can serve heuristic purposes by highlighting the “need for new theory in neglected areas.”¹⁶⁰ However, a number of important methodological questions arise in using disciplined configurative case studies and these are discussed in Chapter 9 on the congruence method.
- *Heuristic* case studies inductively identify new variables, hypotheses, causal mechanisms, and causal paths. “Deviant” or “outlier” cases may be particularly useful for heuristic purposes, as by definition their outcomes are not what traditional theories would anticipate. Also, cases where variables co-vary as expected but are at extremely high or low values may help uncover causal mechanisms.¹⁶¹ Such cases may not allow inferences to wider populations if relationships are nonlinear or involve threshold effects, but limited inferences might be possible if causal mechanisms are identified (just as cancer researchers use high dosages of potential carcinogens to study their effects).
- *Theory testing* case studies assess the validity and scope conditions of single or competing theories. As discussed in Chapter 6, it is important in tests of theories to identify whether the test cases are most-likely, least-likely, or crucial for one or more theories. Testing

may also be devised to identify the scope conditions of theories (the conditions under which they are most- and least-likely to apply).

- *Plausibility probes* are preliminary studies on relatively untested theories and hypotheses to determine whether more intensive and laborious testing is warranted. The term “plausibility probe” should not be used too loosely, as it is not intended to lower the standards of evidence and inference and allow for easy tests on most-likely cases.
- “*Building Block*” studies of particular types or subtypes of a phenomenon identify common patterns or serve a particular kind of heuristic purpose. These studies can be component parts of larger contingent generalizations and typological theories. Some methodologists have criticized single-case studies and studies of cases that do not vary in their dependent variable.¹⁶² However, we argue that single-case studies and “no variance” studies of multiple cases can be useful if they pose “tough tests” for theories or identify alternative causal paths to similar outcomes when equifinality is present.¹⁶³ (See also the more detailed discussion of “building blocks” theory below.)

Researchers should clearly identify which of these six types of theory-building is being undertaken in a given study; readers should not be left to find an answer to this question on their own. The researcher may fail to make it clear, for example, whether the study is an effort at theory testing or merely a plausibility probe. Or the researcher may fail to indicate whether and what kind of “tough test” of the theory is supposedly being conducted.¹⁶⁴

These six research objectives vary in their uses of induction and deduction. Also, a single research design may be able to accomplish more than one purpose—such as heuristic and theory testing goals—as long as it is careful in using evidence and making inferences in ways appropriate to each research objective. For example, while it is not legitimate to derive a theory from a set of data and then claim to test it on the same data, it is

sometimes possible to test a theory on different data, or new or previously unobserved facts, from the same case.¹⁶⁵

Specific questions that need to be addressed in designating the research objectives include:

- What is the phenomenon or type of behavior that is being singled out for examination; that is, what is the class or subclass of events of which the cases will be instances?
- Is the phenomenon to be explained thought to be an empirical universal (i.e., no variation in the dependent variable), so that the research problem is to account for the lack of variation in the outcomes of the cases? Or is the goal to explain an observable variation in the dependent variable?
- What theoretical framework will be employed? Is there an existing theory or rival candidate theories that bear on those aspects of the phenomenon or behavior that are to be explained? If not, what provisional theory or theories will the researcher formulate for the purpose of the study? If provisional theories are lacking, what theory-relevant variables will be considered?
- Which aspects of the existing theory or theories will be singled out for testing, refinement, or elaboration?
- If the research objective is to assess the causal effects or the predictions of a particular theory (or independent variable), is that theory sufficiently specified and operationalized to enable it to make specific predictions, or is it only capable of making probabilistic or indeterminate predictions? What other variables and/or conditions need to be taken into account in assessing its causal effects?

Researchers' initial efforts to formulate research objectives for a study often lack sufficient clarity or are too ambitious. Unless these defects are corrected, the study will lack a clear focus, and it will probably not be possible to design a study to achieve the objectives.

Better results are achieved if the “class” of the phenomenon to be investigated is not defined too broadly. Most successful studies, in fact, have worked with a well-defined, smaller-scope *subclass* of the general phenomenon.¹⁶⁶ Case study researchers often move down the “ladder of generality” to contingent generalizations and the identification of more circumscribed scope conditions of a theory, rather than up toward broader but less precise generalizations.¹⁶⁷

Working with a specified subclass of a general phenomenon is also an effective strategy for theory development. Instead of trying in one study to develop a general theory for an entire phenomenon (e.g., all “military interventions”), the investigator should think instead of formulating a typology of different kinds of interventions and proceed to choose one type or subclass of interventions for study, such as “protracted interventions.” Or the study may focus on interventions by various policy instruments, interventions on behalf of different goals, or interventions in the context of different alliance structures or balances of power. The result of any single circumscribed study will be one part of an overall theory of intervention. Other studies, focusing on different types or subclasses of intervention, will be needed to contribute to the formulation of a general theory of interventions, if that is the broader, more ambitious research program. If the typology of interventions identifies six major kinds of intervention that are deemed to be of theoretical and practical interest, each subtype can be regarded as a candidate for separate study and each study will investigate instances of that subtype.

This approach to theory development is a “building block” procedure. Each block—a study of each subtype—fills a “space” in the overall theory or in a typological theory. In addition, the component provided by each building block is itself a contribution to theory; though its scope is limited, it addresses the important problem or puzzle associated with the type of intervention that led to the selection and formulation of the research objective. Its generalizations are more narrow and contingent than those of the general “covering laws” variety that some hold up as the ideal, but they are also more precise and may involve relations with higher probabilities.¹⁶⁸ In other words, the building block developed for a subtype is self sufficient;

its validity and usefulness do not depend upon the existence of other studies of different subclasses of that general phenomenon.

If an investigator wishes to compare and contrast two or more different types of intervention, the study must be guided by clearly defined puzzles, questions, or problems that may be different from or similar to those of a study of a single subclass. For example, the objective may be to discover under what conditions (and through what paths) Outcome X occurs, and under what conditions (and through what paths) Outcome Y occurs. Alternatively, the objective may be to examine under what conditions Policy A leads to Outcome Y and under what other conditions Policy A leads to Outcome X. Similarly, the focus may be on explaining the outcome of a case or a subclass or type of cases, or it may be on explaining the causal role of a particular independent variable across cases.

Task Two: Developing a Research Strategy: Specification of Variables

In the course of formulating a research objective for the study—which may change during the study—the investigator also develops a *research strategy* for achieving that objective. This requires early formulation of hypotheses and consideration of the elements (conditions, parameters, and variables) to be employed in the analysis of historical cases. Several basic decisions (also subject to change during the study) must be made concerning questions such as the following:

- What exactly and precisely is the dependent (or outcome) variable to be explained or predicted?
- What independent (and intervening) variables comprise the theoretical framework of the study?
- Which of these variables will be held constant (serve as parameters) and which will vary across cases included in the comparison?

The specification of the problem in Task One is closely related to the statement of what exactly the dependent variable will be. If a researcher defines the problem too broadly, he or she risks losing important differences among cases being compared. If a researcher defines the problem too narrowly, this may severely limit the scope and relevance of the study and the comparability of the case findings.¹⁶⁹ As will be noted, the definition of variance in the dependent variable is critical in research design.

In analyzing the phenomenon of “war termination,” for instance, a researcher would specify numerous variables. The investigator would decide whether the dependent (outcome) variable to be explained (or predicted) was merely a cease-fire or a settlement of outstanding issues over which the war had been fought. Variables to be considered in explaining the

success or failure of war termination might include the fighting capabilities and morale of the armed forces, the availability of economic resources for continuing the war, the type and magnitude of pressures from more powerful allies, policymakers' expectation that the original war aim was no longer attainable at all or only at excessive cost, the pressures of pro-war and anti-war opinion at home, and so on. The researcher might choose to focus on the outcome of the dependent variable (e.g., on cases in which efforts to achieve a cease-fire or settlement failed, but adding cases of successful cease-fires or settlements for contrast) to better identify the independent and intervening variables associated with such failures. Alternatively, one might vary the outcome, choosing cases of both successes and failures in order to identify the conditions and variables that seem to account for differences in outcomes.

Alternatively, the research objective may focus not on outcomes of the dependent variable, but on the importance of an independent variable—e.g., war weariness—in shaping outcomes in a number of cases.

We conclude this discussion of Task Two with a brief review of the strengths and weaknesses of the common types of case study research designs in relation to the kinds of research objectives noted above.

First, single case research designs can fall prey to selection bias or over-generalization of results, but all of the six theory-building purposes identified above have been served by studies of single well-selected cases that have avoided or minimized such pitfalls. Obviously, single-case studies rely almost exclusively on within-case methods, process-tracing, and congruence, but they may also make use of counterfactual analysis to posit a control case.¹⁷⁰

For theory testing in single cases, it is imperative that the process-tracing procedure and congruence tests be applied to a wide range of alternative hypotheses that theorists and even participants in the events have proposed, not only to the main hypotheses of greatest interest to the researcher. Otherwise, *left-out variables* may threaten the validity of the research design. Single cases serve the purpose of theory testing particularly well if they are “most-likely,” “least-likely,” or “crucial” cases. Prominent case studies by Arend Lijphart, William Allen, and Peter Gourevitch, for

example, have changed entire research programs by impugning theories that failed to explain their most-likely cases.¹⁷¹

Similarly, studies of single “deviant” cases and of single cases where a variable is at an extreme value can be very useful for heuristic purposes of identifying new theoretical variables or postulating new causal mechanisms. Single-case studies can also serve to reject variables as being necessary or sufficient conditions.¹⁷²

Second, the research objective chosen in a study may require comparison of several cases. There are several comparative research designs. The best known is the method of “controlled comparison”—i.e., the comparison of “most similar” cases which, ideally, are cases that are comparable in all respects except for the independent variable, whose variance may account for the cases having different outcomes on the dependent variable. In other words, such cases occupy neighboring cells in a typology, but only if the typological space is laid out one change in the independent variable at a time. (See Chapter 11 on typological theories.)

As we discuss in Chapter 8 on the comparative method, controlled comparison can be achieved by dividing a single longitudinal case into two—the “before” case and an “after” case that follows a discontinuous change in an important variable. This may provide a control for many factors and is often the most readily available or strongest version of a most-similar case design. This design aims to isolate the difference in the observed outcomes as due to the influence of variance in the single independent variable. Such an inference is weak, however, if the posited causal mechanisms are probabilistic, if significant variables are left out of the comparison, or if other important variables change in value from the “before” to the “after” cases.

However, even when two cases or before-after cases are not perfectly matched, process-tracing can strengthen the comparison by helping to assess whether differences other than those in the main variable of interest might account for the differences in outcomes. Such process-tracing can focus on the standard list of potentially “confounding” variables identified by Donald Campbell and Julian Stanley, including the effects of history,

maturation, testing, instrumentation, regression, selection, and mortality. ¹⁷³ It can also address any idiosyncratic differences between the two

Another comparative design involves “least similar” cases and parallels John Stuart Mill’s method of agreement. ¹⁷⁴ Here, two cases are similar in outcome but differ in all but one independent variable, and the inference might be made that this variable contributes to the invariant outcome. For example, if teenagers are “difficult” in both postindustrial societies and tribal societies, we might infer that their developmental stage, and not their societies or their parents’ child-rearing techniques, account for their difficult natures. Here again, left-out variables can weaken such an inference, as Mill recognized, but process-tracing provides an additional source of evidence for affirming or infirming such inferences.

Another type of comparative study may focus on cases in the same cell of a typology. If these have the same outcome, process-tracing may still reveal different causal paths to that outcome. Conversely, multiple studies of cases with the same level of a manipulable independent variable can establish under what conditions that level of the variable is associated with different outcomes. In either approach, if outcomes differ within the same type or cell, it is necessary to look for left-out variables and perhaps create a new subtype.

Often, it is useful for a community of researchers to study or try to identify cases in all quadrants of a typology. For example, Sherlock Holmes once inferred that a dog that did not bark must have known the person who entered the dog’s house and committed a murder, an inference based on a comparison to dogs that do bark in such circumstances. To fully test such an assertion, we might also want to consider the behavior of non-barking non-dogs on the premises (was there a frightened cat?) and barking non-dogs (such as a parrot). The process of looking at all the types in a typology corresponds with notions of Boolean algebra and those of logical truth tables. ¹⁷⁵ However, it is not necessary for each researcher to address all the cells in a typology, although it is often useful for researchers to offer suggestions for future research on unexamined types or to make comparisons to previously examined types.

Finally, a study that includes many cases may allow for several different types of comparisons. One case may be most similar to another and both may be least similar to a third case. As noted below, case selection is an opportunistic as well as a structured process—researchers should look for whether the addition of one or a few cases to a study might provide useful comparisons or allow inferences on additional types of cases.

Task Three: Case Selection

Many students in the early stages of designing a study indicate that they find it difficult to decide which cases to select. This difficulty usually arises from a failure to specify a research objective that is clearly formulated and not overly ambitious. One should select cases not simply because they are interesting, important, or easily researched using readily available data. Rather, case selection should be an integral part of a good research strategy to achieve well-defined objectives of the study. Hence, the primary criterion for case selection should be relevance to the research objective of the study, whether it includes theory development, theory testing, or heuristic purposes.

Cases should also be selected to provide the kind of control and variation required by the research problem. This requires that the universe or subclass of events be clearly defined so that appropriate cases can be selected. In one type of comparative study, for example, all the cases must be instances of the same subclass. In another type of comparative study that has a different research objective, cases from different subclasses are needed.

Selection of a historical case or cases may be guided by a typology developed from the work in Tasks One and Two. Researchers can be somewhat opportunistic here—they may come across a pair of well-matched before-after cases or a pair of cases that closely fit “most similar” or “least similar” case research designs. They may also come upon cases that have many features of a most- or least-likely case, a crucial case, or a deviant case.

Often researchers begin their inquiry with a theory in search of a test case or a case in search of a theory for which it is a good test.¹⁷⁶ Either approach is viable, provided that care is taken to prevent case selection bias and, if necessary, to study several cases that pose appropriate tests for a candidate theory once one is identified. Often, the researcher might start with a case

that interests her, be drawn to a candidate theory, and then decide that she is more interested in the theory than in the case and conclude that the best way to study the theory is to select several cases that may not include the case with which the inquiry began. Some such iteration is usually necessary—history may not provide the ideal kind of cases to carry out the tests or heuristic studies that a research program most needs at its current stage of development.

Important criticisms have been made of potential flaws in case selection in studies with one or a few cases; such concerns are influenced by the rich experience of statistical methods for analyzing a large-N. David Collier and James Mahoney have taken issue with some widespread concerns about selection bias in small studies; we note four of their observations. ¹⁷⁷ They question the assertion that selection bias in case studies is potentially an even greater problem than is often assumed (that it may not just understate relationships—the standard statistical problem—but may overstate them). They argue that case study designs with no variance in the dependent variable do not inherently represent a selection bias problem. They emphasize that case study researchers sometimes have good reasons to narrow the range of cases studied, particularly to capture heterogeneous causal relations, even if this increases the risk of selection bias. They point out (as have we) that case study researchers rarely “overgeneralize” from their cases; instead, they are frequently careful in providing circumscribed “contingent generalizations” that subsequent researchers should not mistakenly overgeneralize.

Task Four: Describing the Variance in Variables

The way in which variance is described is critical to the usefulness of case analyses in furthering the development of new theories or the assessment or refinement of existing theories. This point needs emphasis because it is often overlooked in designing studies—particularly statistical studies of a large-N. The researcher’s decision about how to describe variance is important for achieving research objectives because the discovery of potential causal relationships may depend on how the variance in these variables is postulated. Basing this decision on *a priori* judgments may be risky and unproductive; the investigator is more likely to develop sensitive ways of describing variance in the variables after he or she has become familiar with how they vary in the historical cases examined. An iterative procedure for determining how best to describe variance is therefore recommended.¹⁷⁸

The variance may in some instances be best described in terms of qualitative types of outcomes. In others, it may be best described in terms of quantitative measures. In either case, one important question is how many categories to establish for the variables. Fewer categories—such as dichotomous variables—are good for parsimony but may lack richness and nuance, while greater numbers of categories gain richness but sacrifice parsimony. The trade-off between parsimony and extreme richness should be determined by considering the purposes of each individual study.

In a study of deterrence, for example, Alexander George and Richard Smoke found it to be inadequate and unproductive to define deterrence outcomes simply as “successes” or “failures.”¹⁷⁹ Instead, their explanations of individual cases of failure enabled them to identify different types of failures. This led to a typology of failures, with each type of failure having a different explanation. This typology allowed George and Smoke to see that deterrence failures exemplified the phenomenon of equifinality. The

result was a more discriminating and policy-relevant explanatory theory for deterrence failures.¹⁸⁰

The differentiation of types can apply to the characterization of independent as well as dependent variables. In attempting to identify conditions associated with the success or failure of efforts to employ a strategy of coercive diplomacy, one set of investigators identified important variants of that strategy.¹⁸¹ In their study, coercive diplomacy was treated as an independent variable. From an analysis of different cases, four types of the coercive diplomacy strategy were identified: the explicit ultimatum, the tacit ultimatum, the “gradual turning of the screw,” and the “try and see” variant. By differentiating the independent variable in this way, it was possible to develop a more discriminating analysis of the effectiveness of coercive diplomacy and to identify some of the factors that favored or handicapped the success of each variant. A very general or undifferentiated depiction of the independent variable would have “washed out” the fact that variants of coercive diplomacy may have different impacts on outcomes, or it might have resulted in ambiguous or invalid results. In addition, the identification of different variants of coercive diplomacy strategy has important implications for the selection of cases.