Quantifying the Qualitative

Information Theory for Comparative Case Analysis

Katya Drozdova • Kurt Taylor Gaubatz

CHAPTER

1

Enhancing Small-*n* Analysis

Information Theory and the Method of Structured-Focused Comparison

This is "the Information Age." The explosion of digitized information has led to important innovations in scholarship across a broad range of fields. Most of these developments have been focused on the emerging computational technologies for efficiently processing and analyzing "big data," the massive streams of digital information generated by our online lives. At the same time, however, appreciation has grown for the role of smaller, contextualized, and more detail-oriented case study techniques. This renewed focus on case studies is due both to the rise of increasingly complex and exclusionary quantitative techniques and to a better understanding of the promise of mixed methods, of working through problems at different levels of focus. Large-*N* and small-*n* analyses are complementary rather than competitive.

In this book, we argue that the core insights about the nature of information that launched the Information Age can be turned toward the task of enhancing small-*n* analysis. We make the case here for the use of information theory as a powerful but accessible tool for making comparative case studies more rigorous and systematic. It is a powerful approach because it provides rigor and replicability without many of the limitations that arise for traditional statistics in situations with severely limited numbers of observations. It is accessible in that it requires only minimal quantitative skills. As we will demonstrate, these techniques can be easily implemented with a simple spreadsheet program. Indeed, our argument will be that "if you can count, you can do it."

Our approach for conducting case studies scientifically draws on the method of structured-focused comparison. The structured-focused method encourages systematic discipline in comparative case studies, but it lacks rigorous tools for assessing the results. The field of information theory provides straightforward quantitative metrics for assessing uncertainty and knowledge gained from and across cases.

Why Quantify the Qualitative? Enhancing Qualitative Analysis With Information Theory

James Gleick (2011) opens his book, *The Information*, with a story about work at Bell Labs in 1948. That year saw the basic research arm of the Bell Telephone monopoly announce two world-changing innovations. The lesser of these, Gleick argues, was the invention of the transistor. The more important was Claude E. Shannon's 1948 paper, "A Mathematical Theory of Communication." That technical paper set out a mathematical foundation for the unified understanding, communication, and measurement of information regardless of its content. Shannon's paper introduced the world to the term *bit* to refer to a discrete unit for measuring information. This shaped the language of ones and zeros now universal and essential to our Information Age. Shannon's paper also included a simple formula to calculate the minimal number of bits required to accurately communicate any information. Humans had pondered information since time immemorial, but there had not been a universal metric until Shannon's information theory-based measures, *information entropy* and *mutual information*.

Shannon's insights were, in the words of Princeton professor Sergio Verdú (1998), "the Magna Carta of the information age." Mathematician Ian Stewart (2012) includes Shannon's information formula among the seventeen most consequential mathematical equations in history. The computer scientist John MacCormick (2013) identifies Shannon's contribution as one of "Nine Algorithms that Changed the Future." Information theory has since grown as a distinct and highly productive branch of applied mathematics and computer science, with implications across many other disciplines, and it is apt for our new application to case study research.

Across a number of fields, the use of comparative case studies is at a watershed moment. The huge advances in big data (itself a direct consequence of Shannon's revolutionary work) have led to increasingly complex and rarefied quantitative methods. At the same time, there is an increasing appreciation for the complementarity of qualitative and quantitative analysis (Johnson, Onwuegbuzie, & Turner, 2007). Qualitative analysis can help validate and contribute nuance to large-*N* analysis as well as generate new hypotheses or rich contextual explanations. The limitation of qualitative analysis is that it has always been dependent on subjective assessment. The analyst collects data on some set of cases and then uses the natural processing power of the human brain to tease out relevant patterns. Information theory provides an exceedingly straightforward and accessible approach that can make qualitative case comparisons more systematic and reproducible. It can make them more scientific.

The core of all science is to infer something valid about the world from the limited data observed. The scientific role of data analysis, including small-*n* case study work, is to discern the information content of the data. Which variables or factors are most informative about an outcome we wish to study? How much can the available data tell us about a phenomenon

of interest? This is essentially an issue of communication—hence the relevance of Claude Shannon's mathematical theory of communication.

In this book, we provide an accessible explanation of Shannon's information concepts and show how they can be effectively applied to make comparative case studies more rigorous and systematic. We demonstrate step-by-step the simple calculation of information metrics to significantly enhance the method of structured-focused case study analysis and the communication of comparative case study results.

Who Needs to Quantify the Qualitative?

This book is for scholars, analysts, and practitioners who need to draw information and insights from relatively modest quantities of data. Many kinds of program analysis and evaluation, for example, might involve small numbers of observations but still require systematic assessment. Policy makers often make consequential decisions based on limited information. Medical researchers interested in the meta-analysis of previously published studies are often challenged to provide clear metrics of aggregation. Scholars across a number of disciplines often conduct more in-depth qualitative study of a set of comparable cases but need to communicate an overarching sense of their findings. In all of these examples and many others, information theory can provide a reproducible metric for systematically understanding the patterns and quality of information contained in qualitative data.

More concretely, the techniques we develop and demonstrate are ideal for working with 4 to 30 comparative cases. For fewer than 4 cases, information theory won't tell you very much that isn't directly observable and will be highly sensitive to changes in any single case. Above about 30 cases, traditional statistical techniques may be as useful, but as we argue, information theory can still provide a powerful measure of uncertainty reduction that does not depend on assumptions about the underlying distribution, is capable of detecting complex relationships that may be overlooked by more traditional central tendency based measures, and can provide a precise measure of independence among variables.

The cases for analysis need to be set up with a single outcome variable and a distinct set of explanatory factors. All of the variables must be amenable to binary coding—that is, to be assigned values of either zero or one. This is straightforward based on systematically measured underlying variables and clear definitions, although developing such definitions to measure complex variables or abstract concepts often explored by qualitative work can be a challenge. The process of clarifying variable definitions and measurements itself carries significant value for case studies independently of information metrics. Cases and conditions that are clearly defined and replicably measured will provide a stronger foundation for

drawing inferences, whether through the techniques we outline in this book or even just with traditional qualitative assessment.

Many areas and kinds of study will fall within these broad parameters and have traditionally been the focus of qualitative work. There are numerous advantages to in-depth qualitative and methodologically careful subjective work. Our analytic approach in no way reduces the ability to draw analytic inferences from comparative case studies in the traditional case study mode. Instead, information theory contributes a systematic and reproducible metric to aid in the analysis and communication of analytic results.

Before delving into these powerful new methods, it is helpful to briefly step back and appreciate the broader context and tradition of policy-relevant scholarship—engaging some of the most critical challenges and advancements of the modern era from the advent of nuclear weapons to the Information Age and beyond—that has produced the structured-focused method as well as motivated our approach to enhance it with information analytics.

Information and Action Under Uncertainty

At the dawn of the Information Age, scholars and practitioners were thinking about information and how it is transmitted and communicated. We are interested in how case studies transmit and communicate information, and so it is worth spending a little time on the essential formative story of how the ideas and concepts that we apply here arose from an overarching agenda of applying fundamental science to vital issues. The origins of information theory stretch to the days before the Internet was invented and computing became ubiquitous, the days when social networking still meant talking directly face-to-face. Since then, information theory has helped transform societies and our daily lives through critical scientific and technological achievements.

In the mid-20th century, many scientific efforts coalesced around winning World War II and the uncertain peace that followed in the Cold War. Nuclear weapons became a terrifying reality that coupled basic science and public policy with the risks of mutually assured destruction. The scientific knowledge that went into harnessing the atom also created opportunities for tremendous advancements in human society through the application of new energy, computing, and intellectual resources for peaceful purposes.

Demand for answers to fundamental research questions arose across the many fields that needed to find better ways to process information and manage uncertainty. Physicists, mathematicians, computer scientists, and engineers as well as social scientists adopted this urgent agenda. Its practical applications ranged from breaking enemy codes and avoiding

a nuclear war to improving communications and creating the new industries that would give rise to the Information Age. Information theory and the developments in case study methodology that we synthesize in this book both emerged to meet these critical challenges.

Origins and Motivations

Alexander George and the Method of Structured-Focused Comparison

The need for decision making under conditions of uncertainty spurred an interest in the systematic use of case studies across a variety of fields. Such work was essential for effective learning from past experience as well as generating basic knowledge.

Alexander L. George, one of the earliest recipients of a MacArthur Foundation "Genius Grant," was an American political scientist and one of the leading theoreticians for the systematic use of comparative case studies in the social sciences. George coined the term *structured, focused comparison* to designate a form of case study analysis that could compete with large-*N* analysis for a claim to systematic and scientific status (George & Bennett, 2005). The method of structured-focused comparison is really just a worked-out version of John Stuart Mill's method of agreement and method of difference (Mill, 1843; Van Evera, 1997). The essential element of structured-focused comparison is the use of theory to clearly and explicitly identify a single outcome variable of interest and a set of causal factors. These variables should be defined and measured consistently across a carefully selected set of cases. The combination of theory with a systematic approach to the cases allows for methodologically rigorous empirical investigation.

George spent most of the 1960s as a social scientist at the RAND Corporation. There, he focused on the challenges of interstate conflict and preventing nuclear war. He developed the method of structured-focused comparison to help glean systematic lessons from sets of theoretically relevant historical episodes.

George sought ways of studying historical cases to draw out valid and policy-relevant lessons. A systematic understanding of the broader phenomena of which the cases were instances would aid policy makers in accurately diagnosing new cases of such phenomena. Policy makers could then make informed judgments about the choice of strategies and actions in new situations, despite the inherent uncertainty of international interactions.

Little was found at the time in the academic literature on methods for the rigorous and systematic study of historical experiences (George & Bennett, 2005). George and his

colleagues set out to devise a case study methodology for conducting comparative analysis in ways that would connect analytical explanations of each case into a broader, more complex theory. This approach aimed to discourage subjective reliance on potentially flawed historical analogies but rather to identify specific causal patterns associated with the alternative outcomes that had resulted from using different strategies under varied sets of conditions.

In order "to analyze past instances of each of the generic problems to identify conditions and procedures that were associated with successful or failed outcomes" (George & Bennett, 2005, pp. x–xi), approaches were developed for converting historical and descriptive explanations of case conditions and outcomes into analytic explanations built from theory-driven variables. The outcome served as a dependent variable to be explained by the conditions analyzed as independent variables. This method did not intend or permit using the findings of a few cases that were not necessarily representative to estimate a probability distribution for the entire universe of instances of, for example, deterrence. Rather, the method enabled what George called "contingent generalizations" intended to help practitioners diagnose new situations and select or prescribe a solution strategy from among the available options, as medical practitioners might do in a clinical setting.

We return to one of Alexander George's structured-focused comparisons in Chapter 4, where we use his landmark study of coercive diplomacy as an example in providing the step-by-step instructions for calculating information metrics. At that same time, we demonstrate the ways in which the information method can enhance this prominent comparative case study.

Claude Shannon and a Mathematical Theory of Communication

Claude E. Shannon's fundamental insights that laid the groundwork for what became known as "information theory" also arose from the national security challenges of World War II and the Cold War. Shannon's work on intelligence ciphers for national defense brought him into contact with Alan Turing, who was in residence at the Bell Labs for two months in early 1943. Shannon conceptualized cryptography and other early computer science applications as instances of general problems of communication across noisy channels. The approach was formalized in his now famous 1948 paper on "A Mathematical Theory of Communication," published in *Bell System Technical Journal*. This paper argued that the fundamental problem of communication was that of "reproducing at one point either exactly or approximately a message selected at another point." The variation of meaning and other factors associated with each message could be systematically related to certain measurable physical or conceptual entities (much like independent variables). The central concern was to reduce uncertainty about the outcome (dependent variable) of a selection of an actual message from a set of possible messages and its delivery to destination intact (success) or not (failure). A systematic

methodology had to be "designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design" (Shannon, 1948, p. 379). This methodology called for a universal perspective on the uncertainty and complexity of information, and that is what Shannon's information theory provided.

Shannon's conception of information had significant implications for a broad class of intelligence problems. Central questions revolved around how much information one needs to solve a particular problem, such as identifying the target and timing of an impending attack, to make an informed decision, such as what actions are needed to stop this attack. How much information is enough for a decision, and how do we know this decision is correct? These insights had practical implications for the collection and interpretation of information intercepted from secret, encrypted, or noisy signals intelligence channels or obtained from human intelligence sources. Intelligence collection is costly and puts people at risk. Questions such as how long to keep an agent in the field or intercept coded signals to decode a message accurately and timely could be decisive. Concrete measures to assess the magnitude of information needed and its uncertainty could constitute lifesaving and potentially strategic game-changing solutions.

In communication theory and engineering, these problems are particular instances of the general phenomenon of information transmission over noisy channels. In other words, how much do we know from the information we have? How much new information do we need to find out what we need to know, and how do we know? How do we assess the chances that this decision is correct? How do we systematically, unambiguously (quantitatively) measure the magnitude of uncertainty reduction or information gain from each new piece of data (such as a new message to be intercepted, new bits of transmission received, new case of a phenomenon to be observed and studied, new factors examined, etc.)? Shannon's information theory answered these questions.

From Cryptography and Communication to Comparative Case Studies

In the analysis of comparative case studies, we are seeking answers to essentially the same questions. Cases are conceptualized as instances of a more general phenomenon being studied. Given that we have the information obtained from the completed case studies—which relate some theorized explanatory factors (independent variables) to uncertain outcomes (dependent variables)—how much do we really know about the outcome, how much more do we learn from analyzing factors and comparing the cases, and how do we know whether we got it right? How do we systematically, unambiguously (quantitatively) measure the uncertainty reduction and the information gain contributed by each case study or by each variable about the uncertain outcome of a phenomenon of which the cases are selected instances?

Shannon's (1948) approach used the concept of "information entropy"—a measure of uncertainty and complexity he adapted from physics. Because the questions Shannon posed tackled fundamental problems of information, the answers had utility far beyond his original applications.

Our application extends the concept of information entropy further into the domain of knowledge gained from analyzing configurations of complex information. The challenge of drawing inferences from comparative case studies can be conceptualized as a problem of information and communication. Knowledge is obtained from information observed and systematically compared based on instances (cases) of a phenomenon. The central question is, *How much does the information contained in the explanatory factors across a set of cases reduce our uncertainty about the outcome of interest?* The same information uncertainty principles that Shannon developed apply.

Our systematic methodology builds on a synthesis of the analytic principles proposed by Alexander George in the structured-focused comparative case method and the rigorous and replicable metrics developed in Claude Shannon's information theory.

Making Qualitative Analysis of Information Systematic: The Method of Structured-Focused Comparison

We live in a world of both too little and too much information. Because there is too little information, we need a scientific approach to help us draw larger inferences from what little information we have. But there is also too much information, in that the world is full of irrelevant noise. This is particularly true in the modern era of big data and ubiquitous Internet access. Identifying useful information can be difficult when it is hidden among the vast quantities of data increasingly being recorded and streamed every second of every day. Identifying accurate information can also be challenging because it requires deeper knowledge and the ability to evaluate information. The solution to these problems is theory combined with appropriate research methods. We use theories to help us identify the outcomes we should be interested in and to identify the factors that might be connected to those outcomes. These theoretical elements allow us to develop hypotheses about the logic of those connections that can then be evaluated empirically.

The method of structured-focused comparison starts from theory. Theory allows us to pick out the most promising explanatory variables. This is particularly important for small-*n* comparative cases where, by definition, we do not have a lot of degrees of freedom. The method then provides a set of systematic guidelines and procedures for drawing valid

inferences and making contingent generalizations as well as cumulating knowledge from comparative case studies.

George and Bennett (2005, pp. 6–7) articulate the method of structured-focused comparison within the broader framework of the role of case studies in theory development in the social sciences. The book codifies the best practices of case studies, clarifies their comparative advantages, and engages the relevant debates in the philosophy of science. The method of structured-focused comparison interacts with all aspects of theory development, from generation of new hypotheses to testing of existing ones. The method's particular strengths lie in engaging typological theories for modeling and explaining complex contingent generalizations. By systematically investigating factor-outcome relationships in different types of case and variable interactions, the method is designed to generate inferences capable of producing "generic knowledge" that policy makers can use. The results can help scholars and practitioners gauge new situations and develop effective strategies informed by theoretically guided empirical insights.

Structured-focused studies often draw together the contributions of many researchers by incorporating multiauthor cases selected explicitly into a common theory-driven design or by creating meta-analytic frameworks where different studies may be compared on relevant elements. These approaches make insights more comparable, which encourages the cumulation of findings from individual efforts into a larger ordered body of knowledge. Complementarities with information-theoretic approaches emerge from the shared focus on systematic and replicable procedures.

Structured-Focused Comparison: What Is It and How to Do It

The method of structured-focused comparison was developed to study historical experience in ways that would generate useful generic knowledge for important policy problems. The method enables drawing explanations of each case of a particular phenomenon into a broader, more complex theory. The systematic results can inform a better understanding of historical events and, most important, help diagnose and deal with possible new cases (George & Bennett, 2005, pp. 67–72).

The logic of structured-focused comparison is simple and straightforward. The method is "structured" around general research questions determined by the investigator to reflect the objectives of the study. Data collection is standardized and guided by asking the same questions of each case. The same replicable procedures are used to collect and analyze the data for each case. These provisions make systematic cross-case comparison and cumulation of the findings possible. The method is "focused" by dealing only with specific, theoretically determined aspects of the examined cases (George & Bennett, 2005, pp. 67–72).

The requirements for structure and focus apply not only to multicase studies but to single case studies as well. Carefully constructed single case studies will enable researchers to replicate the findings and to later join them with other cases in a comparative framework.

What is a "case" in such case studies? The structured-focused comparison method provides a universal definition: *A case is defined as an instance of a class of events* (George & Bennett, 2005, p. 17). The term *class of events* in this context refers to a phenomenon of interest, such as deterrence, coercive diplomacy, wars, peace treaties, terrorist attacks, educational programs, types of economic systems, political regimes, medical treatments, policy decisions, management actions, business startups, or other phenomena to be evaluated, theorized about, and explored though particular cases.

On the basis of the problem chosen for investigation, the researcher establishes the universe—that is, the specific class or subclass of events—from which the cases shall be selected as instances for theoretically driven study. The specific selection of one or several cases is guided by clearly defined research objectives and an appropriate research strategy to achieve that objective. We address the critical role of case selection in Chapter 3. Suffice it for now to emphasize that for valid scientific inference, it is insufficient and inappropriate to choose cases that are simply interesting or well endowed with available data. A logical rationale is required for how a case contributes to investigating some broader phenomenon, of which the chosen case is an instance. Finally, structured-focused case studies identify variables of theoretical interest for purposes of explanation, scientific inference potential, and leverage for policy makers or other practitioners to enable them to influence outcomes (George & Bennett, 2005, p. 69).

Now we can turn to how such case studies are done. We review several examples of structured-focused case studies in the later chapters in the context of our detailed and step-by-step demonstrations of the use of information metrics. The basic process for structured-focused case study is illustrated in Figure 1.1, which also emphasizes where information metrics fit to enhance the results.

As illustrated in Figure 1.1, the implementation process involves three phases, each with several specific tasks (adapted from George & Bennett, 2005, chap. 4–6):

- *Phase I: Theory and Research Design*. Research design tasks apply for all types of systematic, theory-oriented research. The tasks, which are interrelated and should fit together within an integrated design framework, include the following: (1) Specify the problem and research objective. (2) Develop a research strategy to achieve that objective, including specification of the dependent (outcome) variable to be explained or predicted and the independent and intervening variables (explanatory factors) to be explored within the study's theoretical framework. (3) Describe the way the variables vary.

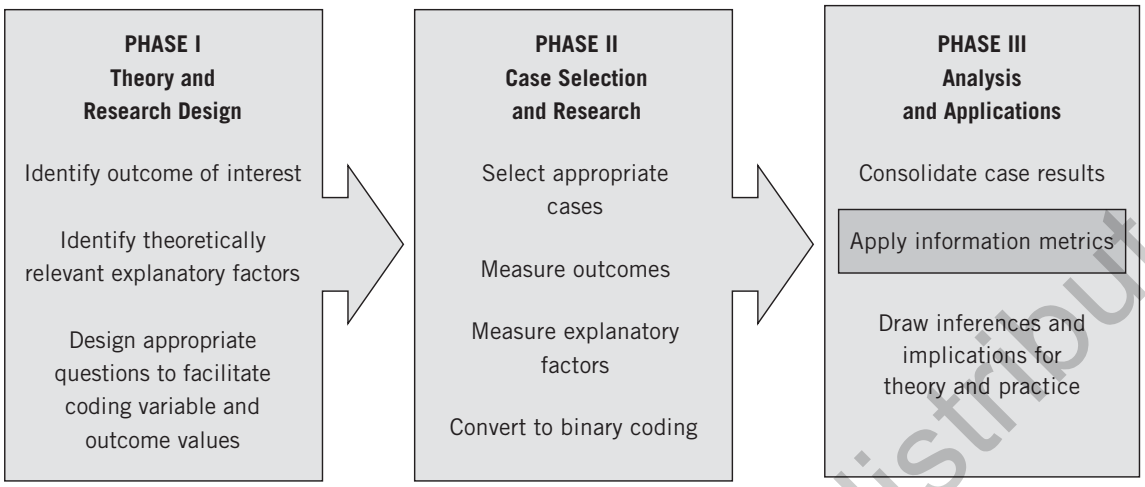

Figure 1.1 The Structured-Focused Comparison Process Enhanced With Information-Theoretic Metrics

Although outcome variables may be classified or coded straightforwardly as success or failure, the underlying phenomenon thus measured may involve much complexity and variability. The researcher needs to understand and reflect this complexity in the coding and measures to maintain external and internal validity as well as replicability of the procedures for comparison and cumulative impact. (4) Formulate data requirements. These are the specific questions to ask of each of the cases (i.e., the questions around which the study is structured and focused) and the standardized procedures so each case study can be carried out replicably, comparably, and consistently with the overall design integrated across all these tasks.

- *Phase II: Case Selection and Research.* Select appropriate cases (Chapter 3 is dedicated to this critical task). The systematic "answers" to the specific research questions for each of the cases constitute the data for structured-focused comparison. The researcher devises provisional explanations and considers the problem of competing explanations and how the evidence can be used to sort out the most informative factors. Descriptive evidence is structured around illuminating the central uncertainties, and all evidence is assessed to develop analytical explanations.
- *Phase III: Analysis and Applications*. Case studies can have implications for theory and practice. Theory development and testing may involve drawing lessons learned or actionable insights for evaluating future cases or decisions. Case study findings may serve to establish, strengthen, or weaken existing explanations, and they may be generalized conditionally to the broader class of phenomena of which the cases are instances or, possibly, to related phenomena. For theory development, results may offer historical

explanation, contingent generalization, and potentially generalization across types. For theory testing, the findings may be used to test competing explanations of cases, test contingent generalizations or the scope or domain of their application, or probe potential implications across types in a typological theory.

The details of these phases and tasks must be adapted to specific investigations, but they provide a set of guidelines rooted in the standards of the scientific method to make structured-focused comparisons qualitatively rigorous and systematic.

The Strengths of Structured-Focused Comparison

As with any scientific methodology, structured-focused comparison has its strengths and limitations. This systematic approach is more amenable to replication and cumulation, and it can be used to develop and assess theory as well as draw contingent generalizations for informing policy decisions.

In a broader methodological landscape, case studies tend to be strong where statistical methods and formal models tend to be weak, insufficient, or inapplicable. This is a basis for the increasingly accepted scholarly recognition of important complementarities in the value and use of alternative methods.

The advantages of structured-focused comparisons for testing hypotheses and developing theory include "their potential for achieving high conceptual validity; their strong procedures for fostering new hypotheses; their value as a useful means to closely examine the hypothesized role of causal mechanisms in the context of individual cases; and their capacity for addressing causal complexity" (George & Bennett, 2005, pp. 19–22).

The Limits of Structured-Focused Comparison

The critical limitation of traditional structured-focused comparisons is that the admirable qualitative rigor of their design and conduct still comes down to relying on subjective assessment of the results. The central problem of many comparative studies is a lack of systematic assessment in the presentation and analysis of cases. We have reviewed case studies across a number of fields and can report that apart from those studies disciplined by *qualitative case analysis* (QCA) methods, which we discuss in Chapter 7, this lack of a comprehensive overview is exceedingly common.

It is also important to note the recurrent trade-offs that are often necessitated by small-*n* research designs. These include trade-offs between the richness of detail and specific explanations of particular events and the parsimony of analytical explanations with inferential

value beyond the immediate cases. Case authors must address the tension between achieving high internal validity for their particular explanations versus making analytical or contingent generalizations for broader classes of phenomena. These trade-offs are often shaped by case selection, as we examine more in depth in Chapter 3.

Finally, George and Bennett (2005) themselves point out limitations in the structured-focused comparison methodology, including a "relative inability to render judgments on the frequency or representativeness of particular cases and a weak capability for estimating the average 'causal effect' of variables for a sample. Potential limitations can include indeterminacy and lack of independence of cases" (p. 22).

Notably, George and Bennett (2005, p. 34) conclude their discussion of their method's limitations with an appeal to scholars to take advantage of the opportunities for multimethod research to correct for limitations and enhance the overall value of this and other methods in our scientific toolkit. Our information theory approach is one answer to this call.

Structured-Focused Comparison: The Bottom Line

As George and Bennett (2005) have argued, the techniques of structured-focused comparison effectively discipline the process of conducting case study analysis. The central limitation of this approach has been the lack of a systematic way to assess the results and frequently weak implementation that fails to build an overarching analytic picture.

We leverage the complementary use of information theory to make structured-focused comparison studies more systematic and rigorous. This approach also forces analysts to present their case study material methodically and to draw clear analytic conclusions. Information analytics are designed to counter the limitation of subjective eyeballing of results with concrete, systematic, and substantive measures that are easy to produce and use. While compensating where possible for case study shortcomings, the information metrics make the information value of results explicit. Most important, while contributing additional tools to enhance small-*n* analysis, this approach preserves the in-depth knowledge and nuance that are the traditional strength of comparative case study research.

Information Theory and Metrics for Qualitative Learning

The information connection for case study research arises from our interest in measuring how informative knowledge of an observed factor is about an outcome. That is, how much does knowing the value of a particular factor reduce our uncertainty about an outcome of

interest? A well-established body of work in information theory provides a solid foundation for producing this kind of assessment in comparative case studies.

Mutual information is a comprehensive measure of uncertainty reduction that is suitable for identifying relationships among variables with the complex or unknown underlying probability distributions that are likely in small-*n* work. It is based on Shannon's information entropy measure, a universal measure of uncertainty of and complexity of information. Traditional statistical measures, such as those based on the central limit theorem or on other sampling dynamics, are not reliable for small numbers of observations. Mutual information is a more sensitive and accurate measure of interdependency among variables because it can uncover relationships not detected by measures based around a central tendency or other limited characteristics, such as correlation, variance, and so on.

The central question may be whether this is the right measure to use. The information metric has a number of advantages: It is not a constructed or sample-based estimate. It is simple to calculate and relatively robust. Because it is not parametric and is simply an expression of the information relationship between an outcome and an independent variable, it is difficult to go wrong in the interpretation of the metric itself. It does not depend on sample size for convergence. It makes no distributional assumptions. Mutual information applies exactly when we cannot verify or rely on such assumptions, as in the world of small-*n* case studies, about highly complex underlying phenomena. It is a precise understanding of the ability of the observed values of one variable to convey information about a second variable, as well as a precise measure of independence when mutual information is zero. The metric is based on an established body of information theory and has acknowledged benefits over other binary correspondence measures or quantitative tools (Brown, Cai, & DasGupta, 2001).

Of course, there are still dangers in making inferences from the case study results. Just as correlation is not causality, information is not causality. But these are the problems of any case study and stem from case selection and interpretation rather than from the information method. Analysts will draw inferences from small-*n* studies. Our argument is that they should do so guided by more concrete metrics rather than by relying only on subjective assessment. Providing a systematic and replicable measure can help make the results of small-*n* studies, including those conducted with QCA, much clearer.

The information metric is attractive conceptually and has comparative advantages that have been thoroughly explored. It is straightforward to calculate and, as we demonstrate in Appendix A, can be done simply in a spreadsheet. The primary mathematical requirement of the analyst is to do counting. If there are fewer than 10 cases, the fingers can be used.

A Road Map for Quantifying the Qualitative

In *Quantifying the Qualitative,* we walk you through the core information-theoretic techniques for the systematic analysis of comparative case studies. We start in Chapter 2 with an overview of information theory and the conceptual groundwork for its use in comparative case study. We also review the basic probability and mathematical concepts (simply based around multiplication and addition) that are used for calculating the information measures.

In Chapter 3, we examine the critical issues surrounding case selection. The information metrics that we develop provide a systematic and reproducible method for describing the relationship between a set of factors and an outcome variable. The ability to draw inferences from the cases to the broader world still depends on the quality of the cases and how they were chosen. One of the virtues of this method is that the metrics themselves are unaffected by case selection issues. They can be applied transparently to convenience samples or even to cherry-picked or biased sets of cases. But the ability to draw inferences from the case studies still relies on the analyst's awareness of case selection issues.

Chapter 4 is where we lay out the step-by-step details for calculating information metrics for comparative case study using a prominent example of a structured-focused study. We show how these measures are calculated from simple data counts and go over their interpretation and their use in communicating systematic analytic results from comparative case study.

In Chapter 5, we provide three diverse real-world examples of information metrics at work. We apply information metrics to reanalyze three published examples of comparative case studies, from the fields of ecology, education, and medicine, respectively. We show how information theory provides replicable metrics for understanding which factors communicate the most information about the outcome of interest. In each case, information theory allows us to draw more detailed and methodical conclusions about these case studies and identify important issues overlooked by subjective assessments alone.

Chapter 6 addresses the issue of sensitivity analysis and confidence intervals for information metrics. Because this approach generates reproducible and transparent quantitative indicators for the relationship between each factor and the outcome, we can easily assess the influence of different kinds of operationalization, measurement, and interpretation errors. We show systematically how dropping or recoding individual cases affects the information conditions. It is also possible to develop confidence intervals on these measures to account for some kinds of error, although we argue that these may be of relatively limited utility for small *n*s.

In Chapter 7, we discuss the relationship of information metrics to the widely used techniques of QCA. This is an approach for reducing the number of factors through the use

of Boolean logic. We show that information metrics can be complementary to QCA and can help better understand the impact of the different reduced-form factors identified through QCA analysis.

Chapter 8 is the traditional conclusion. We review the strengths and limits of the information method for comparative case analysis and discuss some extensions. In particular, we consider the benefits of information metrics for policy makers and policy analysts. Because of its transparency and simplicity, our method can be of great use for the policy community where comparative program analysis is a regular requirement and where the number of cases and quantitative capabilities or interests of the relevant audiences may not allow for traditional statistical analysis or may render it of lesser value.

We have included two appendixes that demonstrate the use of Excel and the statistical software package R for facilitating and automating the calculation of information metrics.

Conclusion

Despite the ongoing revolution in big data, comparative case analysis remains a critical methodology across a broad range of fields. Our purpose in this book is to introduce a simple, systematic, and replicable metric for analyzing the impact of different explanatory variables on some outcomes of interest. This straightforward metric is simple to calculate ("If you can count, you can do it") and thus is accessible to researchers and policy makers with quite varying degrees of quantitative experience or aptitude. Interestingly, it arises from the same foundational work that made big data and the Information Age possible. That foundation is the primary subject of the next chapter.

Additional resources are provided at http://study.sagepub.com/drozdova.