

Each research question is unique and each scholar will be faced with specific methodological dilemmas. I have offered my dilemmas and theoretically guided process of attending to and working within them as an outline in order to indicate the scope of feminist methodological reflection – from research question to publication. At no stage in the research process can the feminist, attentive to power, be confident that her research methodology has adequately interrogated the possibilities for absence, silence, difference, and oppression that the power of knowledge and research can conceal as well as reveal. However, that humility should not obfuscate her responsibility for the choices that she has made. Being ‘in the field’ brings ethical responsibilities. Taking that responsibility requires recognizing that ‘getting it just right’ is a privilege itself, one best shared. For many feminists, theory and empirical work is in the largest sense collaborative (Ackerly and True 2006). Such collaboration requires sharing our dilemmas, our imperfect efforts to work through them, and our partial insights.

4 Case Selection

Audie Klotz

For most researchers, case selection *defines* method: a few cases of a particular phenomenon make a study ‘qualitative’ but a lot of cases turns it into a ‘quantitative’ analysis. Usually a case is equated with a country, and there is often an implicit presumption that some sort of history will be traced. In International Relations (IR), qualitative method typically means a study of one or a few foreign policies, with a decision-making process to be traced at the micro-historical level (George and Bennett 2005). Yet for many questions, say, about globalization, countries are not necessarily the appropriate unit of analysis; economic systems might be. And historical evolution can happen at a higher level of aggregation, such as macro-historical changes in property rights.

Too often, the justification for a research design begins and ends with the rationale for the number of cases, obscuring key issues, such as the unit and level of analysis. In part, this is the result of the problematic conventional dichotomy between qualitative and quantitative methods. Would a project where the researcher uses statistical analysis *within* a single case study be qualitative or quantitative? From a comparative perspective, much of the work on American Politics looks just like that! Intensive analysis of a single case can employ all types of methodological tools without agreement on the degree of general, extensive, knowledge being sought.

Researchers need to remember that cases are cases *of* something. Well-crafted case selection takes into account the universe of possible cases and the logic of comparison implied by the research question. In this chapter, I will show how clarifying the overall purpose of the project and its theoretical framework broadens the rationales for single case studies, paired comparisons, and slightly larger studies. Often I will draw on guidelines by other researchers and suggest their publications for

further reading. Most of my illustrations will come from my dissertation-based book, *Norms in International Relations: The Struggle against Apartheid* (1995), because I can delineate explicitly the sorts of trade-offs and choices that rarely appear in published work.

Cases of what?

Appropriate case selection depends first and foremost on ontology, because any research question relies on core concepts. That brings us to the starting point for case selection: a case of *what*? As Leander (in this book) underscores, questions and concepts remain embedded in theoretical presuppositions. Quite often, these assumptions and subsequent propositions would benefit from clarification. Vagueness is not always the result of sensitivity to context and complexity! What are the key concepts that define a 'case' and what are the key dimensions that should be compared? These are not simply questions of finding indicators, although definitional decisions do hold implications for that stage of research design (Adcock and Collier 2001; Goertz 2006).

When I was formulating my dissertation, for example, I confronted the question of how to conceptualize 'race' in global politics. Was it an ideology? While otherwise quite useful, I found that this standard conceptualization underplayed contestation, and I was intrigued by the international controversy over South African apartheid, particularly the policies of neighboring states. Out went ideology. Alternatively, should I analyze the word race as a linguistic signifier? I found semiotics too focused on specific words, leaving out the material and social dimensions that ideology did capture. Was racism cultural? Yes, in a general sense, but the term 'culture' implied a dense immutability inappropriate for studying IR. I was looking for something less monolithic. Each of these formulations had advantages and disadvantages, but none seemed to capture how the global politics of race appeared to me in the late 1980s.

In the end, I opted to define race in terms of contending 'norms' of racial equality and racial superiority, situating my study in the context of regime theory. Responses to South African apartheid became a *puzzling case of international consensus* that challenged prevailing theories of cooperation based on rational calculations of material interests: Why would racial equality trump domestic jurisdiction? Adcock and Collier refer to this as the shift from a 'background' concept to a 'systematized' one (2001: 530–1; Goertz 2006: 27–57), moving the researcher from abstractions toward measures.

Yet conceptualization is not simply a one-way process, from general to specific. The way the researcher narrows a general concept in order to do empirical research also affects the formulation of the main question – not solely the choice of 'indicators.' Had I defined race through a different theoretical framework, my key question (and subsequent case selection) might also have shifted substantially. More influenced by Michel Foucault and Edward Said, for example, Roxanne Doty (1996) pursued similar questions about race through an analysis of hegemonic discourses, rather than regimes. She queried the constitutive role of race (in the imperial relationships of the United States with the Philippines and of Britain with Kenya), whereas I concentrated on a moment of contestation over it. We started with similar frustrations with the omission of race in IR theories, but our alternative theoretical frameworks led to different key questions, core conceptualizations, and subsequent cases.

Because IR theory lacked any standard conceptualization of 'race' at the time, Doty and I each independently devised a definition to put into practice. That potentially leads to the commonplace critique that case study researchers define concepts idiosyncratically. But the tendency to contextualize concepts need not be an insurmountable problem for comparing across cases. For instance, even an elusive concept like 'regime' has fuelled reams of insightful research on international cooperation and global governance, despite abiding definitional disputes. And applications of constructivism and critical theory have advanced in the two past decades to the point that researchers should be able to find enough common ground in definitions of race (though Doty might not agree with me on this). Most qualitative researchers remain comfortable with a moderately flexible set of characteristics, and many acknowledge the danger of 'stretching' a concept to the point that it loses its essence (Sartori 1970; Collier and Mahon 1993).

Still, case study researchers should avoid undue vagueness and would benefit from the series of questions that Goertz poses (2006: 30–5). He starts with a very basic question: What is the *opposite* of the concept? For example, democracy might be contrasted to authoritarianism or monarchy, depending on the research question. The opposite of racial superiority (of which apartheid was one manifestation) might be non-racialism (not accepting the existence of race as a way to categorize people) or multiracialism (not privileging one race over another). Another useful suggestion is to pay attention to the use of adjectives that modify key nouns, such as 'parliamentary' or 'presidential' democracies, or racial equality versus racial superiority. Since I did not

have Goertz to prod me before I got 'to the field,' I had to figure out some of these distinctions – and their *political* significance – as I parsed the sanctions debates.

How many *possible* cases?

After defining a case and the dimensions of comparison, the researcher needs to decide whether to analyze the full universe of cases or some subset. That requires identifying the 'universe' of cases. It may be small, such as 'world wars' or 'nuclear weapons laboratories' (Gusterson in this book) or large, such as wars or laboratories in general. Clarifying what would be a *non-case* helps for delineating possible cases. A non-case of a nuclear weapons laboratory could be a weapons laboratory or a civilian laboratory. A non-case of war might be violent conflict that remains within the territorial boundaries of one state, a militarized interstate dispute averted, or stable peace. It all depends on the formulation of the research question, although clarifying the universe of cases might, in turn, mean going back to reformulate the core research question.

In my efforts to understand how race affected policies toward South Africa, for instance, I had to decide whether to include all sanctions policies or select a few 'senders' (in the language of that literature). Initially, this task seemed straightforward: list the relevant international organizations and states, then decide the feasibility of including all of them in the analysis. I remain indebted to my dissertation committee for pointing out that only looking at those who adopted sanctions would have prematurely truncated the list of possible cases. My universe of cases quickly expanded, because I needed to include all the *debates* over sanctions to capture times when sanctions *might* have happened but were rejected. (Similarly, see Akerly's discussion, in this book, of her difficulties trying to analyze marginalized and silenced discourses.)

The distinction between cases and non-cases may not be stark. And the gray zone may actually be more interesting (politically as well as theoretically) than the poles – Britain had mixed policies which critics did not consider sanctions at all – but we still need the full spectrum in order to identify its significance. In more formalized terminology, Mahoney and Goertz (2004) offer their 'possibility principle.' They rightly note that most non-cases are actually implausible and the subset of possible cases is much narrower. The trick is to figure out the difference. If you were lucky, like I was, you have a dissertation committee – or a colleague sufficiently informed but not vested in the outcome of the research – to keep you honest about plausibility. Those comfortable with the language of

variables will also benefit from reviewing Mahoney and Goertz's rules of inclusion and exclusion (2004: 657–8).

Some studies go even further to analyze the non-cases, such as policies that are *not* adopted (Price 1997; Tannenwald 2007). Indeed, depending on how the research question is framed, one study's case may be another study's non-case. The use of counter-factual scenarios further expands the range of potential non-cases, although their use raises a whole host of additional issues (Fearon 1991; Tetlock and Belkin 1996). Indeed, I do wonder why post-modernists are derided for challenging the objectivity of historical narratives yet made up counter-factual scenarios can be taken so seriously by social scientists. I prefer to treat these as theoretical formulations or predictions, rather than cases *per se*, because all researchers employ hypothetical 'what if' and 'why not' scenarios, either implicitly or explicitly. And that leads us to the next step in case selection, sorting out the logic of comparison employed in the project.

Which logic of comparison?

Delineating a universe of cases (including non-cases) does not tell a researcher *how* to analyze them, beyond some notion of comparison. Even a single case is not unique, otherwise there is no basis for calling it exceptional. However, the comparison might be against an ideal type. There are diverse ways that researchers parse evidence within a comparative logic. The type of question the study seeks to answer, in turn, depends on its underlying logic.

For example, King, Keohane, and Verba's (1994) controversial advice to increase the number of observations applies a statistical logic of theory testing: the larger the universe (or representative sample), the more persuasive the hypothesized claims about patterns between variables. Yet, for better or worse, redefining a key concept in order to create more observations may fundamentally alter the research question. World wars are not the same thing as militarized interstate disputes. If the research question really is about *worldwide* war, or *nuclear* weapons laboratories, the change may be unwarranted, and a study of only a few events or locations would be appropriate regardless of what the statisticians might think. For questions looking at wars or laboratories in general, though, expanding the study could be beneficial, since it reduces over-generalization from the experiences of great powers or scientists working on secret projects.

Making a choice between, say, five cases that lack the ideal controls versus a near-perfect quasi-experimental paired comparison that shifts

the main question depends on the researcher knowing the embedded logic of comparison. This *sounds* simple. Typically, social science projects seek to make somewhat general causal claims, while those drawing on post-modernism do not. Yet once we start delving into what exactly it means to make a causal claim, neither epistemological position proves to be so obvious.

There is no single 'scientific' logic. Some, inspired by physics, advocate deduction to generate hypotheses, usually followed by statistical testing. Others prefer focused comparison, because it mimics a test tube experiment in a chemistry laboratory. Biology and geology offer other templates. Policy-focused or intensive ethnographic research can also lead to general claims that might be reformulated as hypotheses to test and may convincingly disprove a prevailing theory in particular circumstances (see 'Single Case Studies' below; also Leander and Gusterson in this book). All researchers, therefore, would benefit from clarifying their analytical assumptions by asking themselves the following three general questions.

Does the study seek to test theories?

Avowed social scientists are not the only ones who put forth general claims; a study does not need to use the terms 'hypothesis' or 'variable' to offer theories that can be tested. Furthermore, theory-testing studies are only as good as their hypotheses. And as Ackerly underscores (in this book), there is no guarantee that factors, notably gender (and I would add race), have not been systematically omitted. Framing theoretical insights through the dominant scholarly discourse of testing propositions can lead to productive engagements, even if the initial studies do not use the vocabulary of variables.

For example, critical theorists have raised the visibility of 'omitted variables' such as gender and deserve credit for getting them into the 'equation.' See, for instance, Goertz's overview of how the addition of gender transformed the literature on the welfare state (2006: 88–93). In IR, Foucauldian notions of epistemic power have made in-roads in the past 20 years, as the limits of a materialist conception of power have become increasingly apparent. Not coincidentally, feminist approaches have gained legitimacy along with constructivism (Peterson *et al.* 2005). I adamantly refused to have variables in my dissertation, but these sorts of examples have mellowed me over the years, and therefore I encourage others not to react to scholarship based solely on differences in terminology (Klotz and Lynch 2006).

Are causal claims made in terms of conditions or mechanisms?

One of the most basic lessons about analytical inference is the differentiation between correlation and causation. Its application seems straightforward: use statistical analysis to identify patterns, and then select cases to illustrate which direction the causal effect runs (or figure out if another variable explains both). Case studies trace a causal process that links the proposed independent variable to the dependent one, in order to offer an explanation for why the pattern emerges. This research design has been the bread-and-butter model for mixing methods. More recently, the marriage of rational choice theory (which derives its hypotheses deductively, rather than observing statistical patterns) with historical narrative relies on the same process tracing approach to theory testing.

Elaborating on this as a social *scientific* basis for case studies, advocates of mechanisms have been contributing a lot to the burgeoning literature on qualitative methods (George and Bennett 2005; Gerring 2007). As with all approaches, a focus on mechanisms has its strengths and weaknesses (Checkel in this book). For example, such studies in IR usually demonstrate a chain of decisions by policy makers that presumably links the independent and dependent variables. But because any mechanism can be scaled up or down (Tilly 2001), no one can possibly test all the plausible alternatives. Once again, we are reminded that ontological assumptions about units and levels of analysis are critical.

Yet the problem runs deeper than researchers failing to test propositions about alternative processes. Causal chain narratives downplay contingency and contestation. Indeed, in his (otherwise quite useful) practical guidelines for historical research in IR, Trachtenberg (2006) *advises* writing historical narratives to emphasize the almost-inevitable nature of the outcomes, even if the preceding analysis does take into account alternative scenarios. But these chains of mechanisms, which focus on the presence or absence of factors, are not the only type of causal argument.

Probabilistic claims, articulated in terms of likelihoods, are based on conditions and the conjuncture of various factors at a particular point in time. This latter view shares a contingent quality that post-modernists favor. To see these distinctions in practice, note how Checkel and Dunn (in this book) employ documents and interviews in similar ways to different analytical ends. Checkel aims to create an historical narrative that positions social facts into a coherent story, whereas Dunn offers

a genealogy that highlights contestation over meanings. (Little 1991 offers a succinct overview of different types of causal arguments.)

Are constitutive claims adequately distinguished from causal ones?

My suggestion that probabilistic causal claims may have something in common with genealogy, through their similar emphasis on complex conjunctures, makes it especially important to understand the distinction between a 'constitutive' and a 'causal' argument. In post-modern critiques of social science, these two components often get conflated. Yet even Goertz, a quintessential social scientist, acknowledges the constitutive side of causal theories: 'concepts are theories about ontology' because they are about 'the fundamental constitutive elements of a phenomenon' which play a critical role in explanation (2006: 5).

One might think of this as treating the independent variable in a causal study as the dependent variable in a constitutive one. Foreign policies may 'constitute' identities, by inscribing definitions of Self and Other. Those identities, in turn, narrow the range of conceivable options. Identities thus play a 'causal' role in the sense of making certain choices more likely (and inconceivable ones, extremely unlikely). Formulating constitutive claims as conditional or probabilistic makes them causal *in a particular sense*. This challenges both post-modernists' claims not to be offering explanations and mechanism-oriented social scientists' claims to offer the only true proof of causal connections. Certainly constitutive claims *do not need* to be formulated as causal arguments. But I do think it helps to avoid tautology, which is particularly prevalent in arguments about the 'mutual constitution' of structures and agents (also see Hoffmann in this book).

There is at least one other advantage of *trying* to think of constitutive claims in causal terms: it encourages the researcher to think about what it means for a proposition to be 'wrong.' I do not mean 'falsification' in the narrow sense of largely discredited positivist standards for refuting theories. By 'wrong,' I mean that researchers benefit from thinking about what sorts of evidence might make a particular claim untenable. For example, the role of identity in foreign policy is a common theme in critical security studies. Quite likely, identity is defined in a way that precludes the possibility that there is *no* identity. ('What is the opposite of the concept?') But the question could be reformulated to *ask whether* specific interests conflict with specific identities, *rather than presuming* that the former derives from the latter.

Similarly, if the researcher assumes *multiple* identities, then the question might revolve around what a *dominant* identity might look like

('What adjectives modify the concept?'). Evidence might lead to the conclusion that a particular identity is *not* dominant; it might be subordinate (Neumann in this book), marginalized (Dunn in this book), or silenced (Ackerly in this book). Similarly, if the boundaries of identities are posited to be fluid rather than fixed, another common assumption, what would a relatively stable identity look like (Hermann in this book)?

My point is that researchers from many positivist and post-modern perspectives do parse evidence along similar lines, despite dissimilar philosophical moorings. This suggests potential for complementary insights, if researchers are willing to focus on their logics rather than labels. Recognizing this, we can put pluralism into practice in the selection of cases.

Single case studies

Often single case studies emerge out of an empirical puzzle. We see something that does not fit our expectations based on prevailing theories or conventional wisdom. A researcher, already knowledgeable about a part of the world or particular issue, may have some hunches about what is happening and perhaps some critiques of dominant frameworks. For example, my dissertation built on the observation that materialist theories were of little use for understanding why Zimbabwe risked so much to condemn racial segregation in South Africa. Simply put, apartheid should not have been an international issue if the Realist building blocks of IR, such as sovereignty and balance of power, were accurate. That observation, however, did not tell me what an alternative theory might be, nor did it tell me whether this one anomaly justified the wholesale rejection of Realism.

As my research question emerged, I readily found theoretical arguments that offered a plausible alternative – ideational – framework in Kratochwil and Ruggie's (1986) critique of regime theory. However, this nascent constructivism did not offer a specific theory to test. Indeed, it resisted the whole endeavor of testing theories in the conventional sense! My research into why states and international organizations censured South Africa became a 'plausibility probe' to see if meta-theoretical arguments about the constitution of interests could be translated into empirical research. I presented Realism and Marxism as materialist foils to highlight key aspects of my alternative ideational framework, but the study itself was not designed as a *test* of any theory (evident to anyone who ventures back to read the dissertation's theory chapter).

Plausibility probes are certainly not the only option. Single case studies can be used effectively to test theories if they fall into one of two categories. Some cases should be 'easy' for a theory to explain, yet it falls short. Others are 'unlikely' for a theory to explain, yet it does surprisingly well. There are various labels out there but these are the two general logics. Not all single case studies will fall into one of these two categories. Just the opposite: rarely will such a crucial case be available, but its analytical usefulness can outweigh many large-*n* studies.

An easy case can readily be confused with a plausibility probe, but the distinction is significant because each relies on the opposite logic. The exact same empirical evidence can contain more than one theoretical implication, but not all are of equal significance. For example, imagine that the evidence I gathered did show that norms could reasonably be interpreted as justifications for the pursuit of deeper material interests. An ideational approach would not be a better explanation of the censure of South Africa for racial discrimination – there would be little reason to reject Realism in favor of a new (and barely formulated) alternative theory. Yet a conclusion that Realism indeed could explain the putative weakness of norms would also not, in any strong sense, confirm the theory because South Africa was not a 'great power' (among other issues). Simply put, apartheid was a trivial case for testing Realism.

The second type of theory testing based on a single case is a 'least likely' scenario. Again, this should be distinguished from a plausibility probe, because the two may look similar. Unlike the easy case, a plausibility probe may follow the same logic as a hard case; the difference is the relationship between the theory and the empirical evidence. For example, if constructivist theory had been articulated in a less meta-theoretical way when I plunged into my dissertation, I might have framed it as a 'least likely' study because of the substantial amount of evidence in favor of materialist arguments (strategic resources in southern Africa, markets, and such). Other studies around the same time did directly target those theories by focusing on actors and arenas that prevailing theories considered most important: the World Bank (Finnemore 1996) rather than the Commonwealth, for instance (also the contributions in Katzenstein 1996). Not coincidentally, it was the part of my study on the *United States* that got published in *International Organization* (IO).

The value of single cases – perhaps more so than other selection rationales – depends in particular on the *status* of the theory that underpins it. In the late 1980s, constructivism had not been articulated to the point where, epistemological disputes aside, it could have been

tested – those 'importing' frameworks from other disciplines may face a similar situation. When I revised my dissertation for publication, I sought in the conclusion to translate my framework into more detailed claims that others could subsequently probe. In retrospect, I might have formulated these suggestions in terms of interests or identities as the 'dependent variable,' had I felt more comfortable with that vocabulary.

Now, if someone were to do a similar study, I would expect to see a research design that is built on the logic of easy or least likely cases, because the basic insight that norms or identities 'matter' is no longer novel. With the plausibility of the theoretical claim established, the value of doing additional single case studies (aside from the inherent value of knowing more about a particular place or issue) diminishes. As the circumstances that warrant the use of a 'crucial' case are limited, the research design questions shift to consider carefully paired comparisons or a larger set of cases instead.

Paired comparisons

Experimental logic makes carefully paired comparisons most acceptable for 'positivists' who aspire to test hypotheses. Yet given the infinite number of hypothetically possible variables across diverse levels of analysis, even carefully paired comparisons are inevitably easy to challenge. Outside the laboratory, as social scientists readily admit, ideal conditions will rarely exist. I refrain from saying 'never,' in recognition of a growing interest in field experiments, but these would be tough to apply widely in the IR context. Simulations, either with people or with computers, also offer potential insights, but they remain heuristics (Hoffmann in this book).

The closest approximation is the exploration of a single case over time, sometimes called 'within case' comparison, because it enables a researcher to hold many potential variables relatively constant. What might initially appear as one case turns into a comparative study. The best way to make this longitudinal approach work is when an 'exogenous shock' – the dramatic shift in an independent variable – enables the researcher to track closely what else does and does not change. Otherwise, there is nothing truly *paired* about breaking one case into component parts over an extended period of time. Simply 'tracing' the 'history' of a single case over time does not really take the logic of comparison seriously.

Yet even when there is evidence of such a sharp break, establishing historical stages remains difficult. Questions include how far back to

go and on what criteria to demarcate eras (see Neumann in this book). Quickly a comparison between two periods devolves into a longer study. And then researchers need to look beyond 'pairing' to tackle the difficulties of slightly larger small-*n* studies – the sorts of problems that George's notion of 'structured focused comparison' sought to alleviate (George 1979; George and Bennett 2005). I am not suggesting that we should abandon cross-temporal studies – we simply should not treat them as a special form of rigorous comparison.

The limitation which I find less frequently acknowledged in otherwise sensible discussions is whether paired cases are truly independent units or events. Tilly's (1984) notion of 'encompassing' comparisons comes close; some people use the phrase 'world time' to denote the importance of shared global historical context. Given increasing emphasis on 'globalization' across the social sciences, the question of inter-connections between cases (and not just variables) needs to be confronted explicitly. If countries, the most common unit of analysis, are not independent, then researchers need to figure out ways to control for external factors that may not appear as variables in the relevant literature. For instance, globalization has produced a new interest among comparative politics specialists in norms that diffuse to the local level, and they increasingly acknowledge significant cross-case interactions, such as emulation.

In this context, my South Africa study might be viewed as a study of the evolution of a particular norm (anti-racism). If there were a critical juncture, at which point one could claim that the norm emerged or consolidated, then a 'before' and 'after' study could be treated as a paired comparison, with most key variables either held constant or at least readily identified. In the United Nations debates over apartheid, 1963 marks such a turning point: the Security Council rejected a domestic jurisdiction defense in favor of a 'threat to peace and security' argument. However, in the Commonwealth, 1961 marks the key break: South Africa declared republican status and withdrew from the organization. We cannot do a structured, focused, *paired* comparison across these two organizations, because we cannot apply the same timeframe. Furthermore, the UN decisions took place in the context of prior Commonwealth debates, while those took place following earlier challenges to domestic jurisdiction by India going even farther back than the UN founding. We can learn a lot about, say, majority voting versus consensus by pairing these organizations, but they do not offer independent cases.

With few single cases passing muster as 'crucial' and so many inherent problems in paired comparisons, most qualitative studies fall into the

murky range of small-*n* studies. This gray-zone of 'more than two' but 'less than whatever is statistically significant' presents difficult terrain for case selection. The demands for detailed evidence garnered by a solo researcher are still possible but the results are inevitably more superficial. Some of these constraints can be alleviated through collaborative projects, but dissertation writers are less likely to gain funding for that, unless they work as part of a supervisor's larger project. Therefore, I assume that the trade-offs are faced by individuals.

More-than-Two but Not-a-Lot

Clearly, there is no single formula for dealing with multiple cases. Particularly for studies that start with an empirical puzzle, rather than a theory to test, some of the parameters of case selection are dictated by social realities and historical circumstances (see Dunn in this book). Yet I fear that research designs too often reflect the typical structure of a book: a magic trinity of three case study chapters, along with an introduction and conclusion, comprise a readable and reasonably priced volume. My goal in this section, therefore, is to get away from that trinity without tossing out the possibility that three case studies may indeed be appropriate.

For instance, in my study of international reactions to apartheid, I *could* have analyzed a wide range of international organizations and foreign policies. Yet it made sense to focus on the three communities in which South Africa had historically played a role: the international community (represented by the United Nations), the Commonwealth (initially as a Dominion within the British Empire but then during decolonization), and Africa (as a result primarily of geography rather than choice). Within each of these three communities, I analyzed the collective decisions of an international organization (UN, Commonwealth, and Organization of African Unity) and a key state within each group (the United States, Britain, and Zimbabwe). The result was, indeed, a reasonably priced book that I have been told is readable, as well as fairly convincing to specialists of each of these communities. I cannot complain too much about the magic trinity. But the choice of three communities was primarily inductive, the result of the historical legacies of South Africa's origins as a state in the international system.

Yet, on closer examination, counting the cases in my study is tough. These communities, taken together, comprise a single case of international cooperation (to condemn institutionalized racism in South Africa). In that sense, these are *not* six independent cases of sanctions

policies (in three instances across two types of actors). But if the theoretical focus were theories of decision-making, rather than regimes, they could be. In another sense, each pair of organizations and states could count as a case for tracing norm diffusion (but note the problem of interconnections discussed above). Going that route, the effects of sanctions on South Africa should also be considered a case, but it would not be a 'structured focused comparison,' because South Africa was the target while the other countries were senders of sanctions.

What is a confused researcher to do? My general advice, and my constant refrain in this chapter, is to remain mindful of the theoretical framework and core question, which do lead to reasonable conclusions about relevant cases at appropriate levels of analysis. And do not flee back to the world of single cases studies because they seem simpler; most of them lack analytical leverage. Yet that leaves my students profoundly unsatisfied, and I confess that I too remain uneasy. Other researchers (who also, notably, teach methods) have offered two directions for honing the selection of multiple cases: typologies and fuzzy sets. I am not yet convinced by either but both deserve serious attention from anyone trying to sort out this gray zone of Not-a-Lot of cases.

Typologies provide a fruitful path between the extremes of unattainable universal generalization and idiosyncratic contextualization. One of the advantages of a typology is that it offers an escape from the search for a crucial case or an elusive paired comparison by offering the possibility of comparing one or more cases against an ideal. Think about the adjectives often attached to concepts like democracy or war. These can easily be turned into descriptive or analytical typologies that differentiate forms of a phenomenon. And these typologies can be linked to constitutive or causal claims. One might explore a number of cases to illustrate the full range or concentrate on one cell, depending on the research question. (For elaboration and advocacy of 'typological theories,' see George and Bennett 2005: 233-53.) So far, so good.

Still, I advise caution, because it is seductively easy to draw up a two-by-two table for just about anything. That leads to a tendency to construct a dubious typology that justifies research that you already know you want to do or, especially for seasoned scholars, that relies heavily on research you have already done. I did just that for my dissertation: drawing on the sparse literature on pariah states, I identified two descriptive factors, which were only evident together in the South Africa case. I intended eventually to examine other historical examples (since

I claimed that none of the other contemporary cases were comparable). Although I dropped this convoluted foray into typologies for the book, I hesitate to dismiss the exercise completely. It prodded me to think more historically, which I generally was not inclined to do. And my favorite comparison, the Confederacy during the American civil war, inspired me to write a spin-off article comparing the abolitionists with the anti-apartheid activists (Klotz 2002). In the end, I still think that *insightful* typologies can help us avoid some of the difficulties of comparisons.

Another option gaining followers especially among those seeking to bridge quantitative and qualitative studies is Ragin's notion of 'fuzzy sets,' also known as 'qualitative comparative analysis' (Ragin 2000). Since his approach is full of technical terminology, I will simply mention here a few of the overarching goals that might encourage skeptics to take an initial look at some of his guidelines.

Rather than force complex concepts into rigid conceptual boxes, the notion of fuzzy sets accepts some inherent ambiguity. Concepts comprise a cluster of key characteristics, but no single feature is essential. Thus a 'case' of something includes some, but not necessarily all, of these core dimensions. Here is where the logic of Boolean algebra comes into play, and along with it, specialized terminology and formal notation. Anyone familiar with on-line library searches knows that typing 'and' gets a smaller number of hits than 'or' - it is the same logic. One might link this to typologies, for instance, by defining ideal types in terms of the most exclusive features ('and') while recognizing that cases will evince a subset ('or') of those characteristics (Goertz 2006: 84).

I find this logic appealing, because it helps me wrestle with a basic empirical question: should South Africa be considered a democracy? By the standards of the late 1700s, it certainly should - show me any political system based on universal suffrage at that time! By the standards of the late 1900s, its parliamentary elections without adequate representation clearly did not satisfy most definitions. Since electoral dynamics among white votes did play a significant role in the transition to inclusive democracy, the existence of certain features of democracy should not be overlooked. Also, there would not be much to the democratic peace literature if we used universal suffrage as a necessary feature for defining democracies in earlier times. Fuzzy sets move researchers away from essentialist terminology - which is also a major goal of constructivism and critical theory. Whether it can deliver on this potential for building conceptual bridges without getting mired in the jargon of its formal notation remains to be seen.

Conclusion

Perhaps because I grew up in a family of chemists, I never doubted the value of case studies. Paired comparisons come closest to controlled test tube experiments where one chemical agent (potentially) alters a reaction. No correlation will offer anything as compelling in terms of *causal* inference. Yet, I also never had illusions about the practice of science. For me, the laboratory was not the idealized space that philosophers contemplate; it emitted a distinctive aroma and was populated by human beings. And sometimes scientists – just like any other humans – have been extraordinarily successful in propagating ideas that subsequently appear quite ridiculous (Klotz 1986). I have never expected social reality to mimic molecules, because people are not objects. I also appreciate that scientists, like ethnographers, find some of their greatest insights while looking for something else.

Qualitative researchers of my generation had little to offer in terms of a methodological rationale. Scholars oriented toward theory testing easily dismissed our single case studies as 'thick description,' caricaturing Geertz's famous 1973 essay of that name. I distinctly recall an awkward job interview situation that followed along these lines. Fortunately, we have come a long way in the past 20 years. The significance of single case studies for theory testing is still debated, but it is better understood. And its significance for theory building is widely accepted. Greater attention can now be paid to the messy middle of more than two but less than whatever is statistically relevant. Let me reiterate that these are research design questions that barely begin to address subsequent methodological questions of how to do the actual empirical study within the cases. The remaining chapters in this book do that exceptionally well.

Acknowledgment

Wagaki Mwangi and Deepa Prakash provided insightful written feedback on a short turnaround, for which I am especially grateful.