The Challenge of Conceptual Stretching in Mixed-Methods Research

Ariel I. Ahram

University of Oklahoma

(arielahram@ou.edu)
Editor
Andreas Schedler (CIDE, Mexico City)

Editorial Board
José Antonio Cheibub, University of Illinois at Urbana-Champaign
David Collier, University of California, Berkeley
Michael Coppedge, University of Notre Dame
John Gerring, Boston University
Russell Hardin, New York University
Evelyne Huber, University of North Carolina at Chapel Hill
James Johnson, University of Rochester
Gary King, Harvard University
Bernhard Kittel, University of Oldenburg
James Mahoney, Brown University
Cas Mudde, University of Antwerp
Gerardo L. Munck, University of Southern California, Los Angeles
Guillermo O’Donnell, University of Notre Dame
Amy Poteete, Concordia University, Montreal
Frederic C. Schaffer, University of Massachusetts, Amherst
Ian Shapiro, Yale University
Kathleen Thelen, Northwestern University

The C&M working paper series are published by the Committee on Concepts and Methods (C&M), the Research Committee No. 1 of the International Political Science Association (IPSA), hosted at CIDE in Mexico City. C&M working papers are meant to share work in progress in a timely way before formal publication. Authors bear full responsibility for the content of their contributions. All rights reserved.

The Committee on Concepts and Methods (C&M) promotes conceptual and methodological discussion in political science. It provides a forum of debate between methodological schools who otherwise tend to conduct their deliberations on separate tables. It publishes two series of working papers: “Political Concepts” and “Political Methodology.”

Political Concepts contains work of excellence on political concepts and political language. It seeks to include innovative contributions to concept analysis, language usage, concept operationalization, and measurement.

Political Methodology contains work of excellence on methods and methodology in the study of politics. It invites innovative work on fundamental questions of research design, the construction and evaluation of empirical evidence, theory building and theory testing. The series welcomes, and hopes to foster, contributions that cut across conventional methodological divides, as between quantitative and qualitative methods, or between interpretative and observational approaches.

Submissions. All papers are subject to review by either a member of the Editorial Board or an external reviewer. Only English-language papers can be admitted. Authors interested in including their work in the C&M Series may seek initial endorsement by one editorial board member. Alternatively, they may send their paper to workingpapers@concepts-methods.org.

The C&M webpage offers full access to past working papers. It also permits readers to comment on the papers.

www.concepts-methods.org
There is today widespread enthusiasm today for mixed method research (MMR) in social science. In their seminal *Designing Social Inquiry*, King, Keohane and Verba (1994, 1995) expound on a single unified logic of inference for qualitative and quantitative research which makes it possible to combine the two effectively. Brady and Collier (2004) reiterate the mantra of “shared standards” between qualitative and quantitative work and invite collaboration between them. Gerring (2007) argues emphatically for the equivalence and logical consistency between case and statistical research methods and some go so far as to question the need to maintain the distinction between qualitative and quantitative work (Levy 2007; Prakash & Kottie 2007). In 2006, the American Political Science Association’s Qualitative Methods Section voted to change its name to include mixed methods, and an interdisciplinary handbook and journal devoted to MMR have also recently appeared.¹ These efforts have done a great deal to cut through the now stale meta-theoretical debates between the presumed “subjectivism” of qualitative research and “positivism” of quantitative research in social science (Lin 1998).

Despite advances at the theoretical level, though, many questions remain about how methods are mixed in practice. This paper takes a less sanguine view in addressing

¹ These are, respectively, Tashakkori & Teddlie (2002) and *The Journal of Mixed Methods Research*. 

1
a hitherto little noticed practical limitation of MMR: the inevitable “stretching” of concepts as they are translated and transported between qualitative and quantitative usages (Sartori 1970). It argues that the impact of conceptual stretching depends on the research strategy employed, specifically whether or not MMR seeks to triangulate between methods and verify or validate the findings of one against the findings in other method (Bryman 2004). Stretching is an Achilles heel in any claim that MMR findings are mutually validating. Simply because qualitative and quantitative findings “point in the same direct”—statistical significance and coefficient signs match the outcome of a case study—does not make them any more likely to be true, since the concepts applied in one methodological component are not equivalent to those applied in the other. It is impossible for qualitative and quantitative methods to say they same thing because they are talking about different things. Still, this paper shows that conceptual stretching can be a source of inferential leverage if studies seek instead to elaborate or enhance theories of social change by gaining complementary perspectives of both correlational patterns and mechanistic processes. The results are not necessarily more valid, but more comprehensive. Ultimately, MMR hinges on devising qualitative and quantitative concepts that are close enough to be commensurate but disparate enough that the benefits of both conceptual forms are brought to bear on the same puzzle.

‘Thick’ and ‘Thin’ Concepts Contrasted

The choice between qualitative and quantitative methods in social science is commonly presented as a trade-off between different yet equally valid methodological toolkits (Anderson 2005). Case-based, qualitative methodology begins with an
examination of outcomes in a small number of specific instances and works backwards to identify causal factors. The approach to explanation is to study the “causes of effects” of particular cases using narrative process tracing, most-similar and most-different designs, and other quasi-Boolean logics. The result is a complex etiological model, but one bounded by the scope of conditions used to control for the initial selection of cases. In comparison, variable-based, quantitative methodology begins with the “effects of causes,” estimating average effects of independent variables across a wide number of fairly homogenous cases using regression analysis and other statistical tools. Models of causation are presented as abstract algebraic equations, whose relevance is bounded by assumptions of unit homogeneity, independence, and linearity of effects (Mahoney & Goertz 2006).

Adcock and Collier (2001) and Gerring (2001) have gone the furthest to identify common principles that bridge qualitative and quantitative approaches to conceptualization. First, in both qualitative and quantitative methods, conceptualization is an iterative process of deductive and induction, linking the basic conceptual notions and empirical indicators. Concepts are built using deductive theorizing about the relationship between each level and are then modified inductively by testing the ‘fit’ between indicators and data. Second, in both qualitative and quantitative methods, concepts are laden with causal hypotheses. Whether used as independent or dependent variables, it is the causal power of concepts that make concept-formation such an important part of the social science methodology. Third, all concepts possess two alternative dimensions of intension—systematic and explicit definition at the secondary level—and extensions—the range of cases which fall under the concept. When intension
is described through necessary and sufficient conditionality (e.g., only a country that hold
elections AND protects civil liberties is considered a democracy), intension and extension
are inversely related; the act of adding criteria decreases the number of cases to which the
concept can be applied. When intension is described through family resemblance and
substitutability principles (e.g., a country that provides old age pensions OR workman’s
compensation OR health insurance is considered to be a welfare state), intension and
extension are directly related.

Table 1: Comparison of Quantitative and Qualitative Conceptualizations

<table>
<thead>
<tr>
<th></th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intension</strong></td>
<td><strong>Low.</strong> Conceptual definitions are simple.</td>
<td><strong>High.</strong> Conceptual definitions are complex.</td>
</tr>
<tr>
<td><strong>Extension</strong></td>
<td><strong>High.</strong> Numerous cases can fit under a single conceptual rubric.</td>
<td><strong>Low.</strong> Small number of cases can fit under a single conceptual rubric.</td>
</tr>
<tr>
<td><strong>Causal Properties</strong></td>
<td>Latent and internalized.</td>
<td>Embedded and externalized.</td>
</tr>
<tr>
<td><strong>Measurement</strong></td>
<td><strong>Scoring</strong> of interval scales based on observation of indicators.</td>
<td><strong>Categorization</strong> of nominal and ordinal categories based on observed ontological properties.</td>
</tr>
<tr>
<td><strong>Error Management</strong></td>
<td>Error is assumed as non-systematic.</td>
<td>Error is minimized by conceptual reconfiguration.</td>
</tr>
</tbody>
</table>

But shared standards do not make equivalent, much less identical, analytical
techniques, as shown in Table 1. Coppedge (1999) eloquently describes the tension
qualitative and quantitative concepts and its impact on etiological explanations. There
are two important facets to this distinction: first, the degree of intension versus extension
in concepts; second, the location of causal properties within conceptual definitions and
variable operations.

Quantitative researchers sacrifice definitional complexity in order to gain a
significant sample size, collapsing secondary level characteristics entirely and focus on
relatively simple indicators. Their concepts are ‘thin but broad,’ because they have a high extension (incorporate a large number of cases) and low intension (maintain a relatively simple conceptual definition). Concept systemization is a matter of the additive accumulation of indicators of a latent quantity, and variable measurement becomes basically a function of numerical *scoring* on a ratio scale. By comparison, qualitative concepts are thick but narrow. Conceptual definition are complex, involving combinations of necessary/sufficient conditions and substitutable/family semblances. Operationalization is a function of *categorization* of cases into ontological classes. The methodological goal is to fit a small number of cases accurately under a conceptual rubric in such a way as to eliminate any measurement error.

The different approaches to conceptualization also lead to different forms of variable manipulation in hypothesis testing, as illustrated in Diagram 1 and 2 below. For quantitative variables, causal properties are internalized of causal properties and represent accumulations of indicators of latent effects. Conceptual definitions are typically too thin to offer an explanation, then, of how these concepts could cause changes in other variables. Statistical techniques, therefore, focus on identifying consistent correlational patterns, without imputing causal properties to these relationships. On the other hand, qualitative variables externalize their causal hypotheses, focusing on the how one variable might influence other variables. While qualitative methods cannot identify broadly consistent correlational pattern, they can form the building blocks for causal hypotheses about processes of social change (Goertz 2006: 55-65; Mahoney 2003; King, Keohane, & Verba 1994: 151-7).
The differences between qualitative and quantitative concepts are readily grasped using two commonly-cited variables in social science as examples: development and regime-type. Quantitative research can make ready use of per capita GDP as a
continuous variable to measure development, since GDP growth is a result or “symptom” of development. Other factors, such as per capita electricity usage or the size of the industrial workforce can be treated as complimentary indicators of this latent concept. Since the variable is expressed in ratio format, one can easily tell, for instance, that the United Kingdom (at $35,000 per capita GDP) is five times more developed than Ukraine ($7,000) while virtually equivalent in development to Germany and Finland. Qualitative researchers, in contrast, cannot make use of these kinds of fine-tuned distinctions. Instead, they rely on a categorical set that treats certain degrees of difference as less relevant than others. In the same example, the United Kingdom, Germany and Finland are categorized together as fully or highly developed, while Ukraine is moderately developed (Ragin 2000).

On the other hand, regime-type is a concept resistant to quantification because of its multidimensionality and manifest types and sub-types. It is difficult to maintain, as is necessary in quantitative analysis, that democracy causes elections; rather, elections are a property of democratic regimes. Other attributes, such as the state of civil-military relations and the protection of civil rights, are also important features of different regime-type categories. Case-based researchers can easily apply a proliferation of such classifications and typologies (Linz 2000; Collier & Levitsky 1997). In fact, many conceptualizations of regime-type are tailored to specific regions (Van de Walle & Bratton 1994). By comparison, quantitative analysis relies on numerical data sets, like Freedom House and POLITY. At best, these datasets reveal that two countries are equally democratic (or undemocratic). On quantitative scales, Iran is equivalent to Swaziland, since both received a Freedom House Ratings of “6” for 2008. Even though
the values are expressed in ratio format, there is no substantive meaning to the distance between intervals, leaving hazy meaning of the difference between Iran and Armenia (Freedom House Score of “5”) (Munck & Verkuilen 2002; Hanson & Kopstein 2005).

Examining the practical application of conceptualization also demonstrates that the intension of a concept is almost always inversely related to extension. The simplicity of conceptualization in quantitative research is premised on the assumption of unity homogeneity. The more elaborate the conceptualization and complex the secondary-level definition, the more difficult it becomes to operationalize a measure that effectively summarizes a large sample and the more tenous this assumption becomes (Gerring 2007: 52). By contrast, qualitative researchers respond to substitutability criteria by pursuing contextualized comparisons, in which different but equivalent features across a relatively small number of cases are compared (Locke & Thelen 1995).

Qualitative scholars commonly see statistical variables as vacuous terms far removed from the original ‘thick’ concept they claim to represent. Quantitative scholars retort that qualitative scholars use one concept per case, creating studies so idiographic as to prevent theories and concepts from ever traveling or being tested again. Ultimately, though, the different techniques correspond with the different overall approaches to explanation used in qualitative and quantitative research. Conceptual thickness is necessary to account for specific outcomes of a particular case, while conceptual breadth is required when the goal is elucidate general patterns within a large homogenous population of cases (Mahoney 2003).
Conceptual Stretching in Theory and Practice

MMR involves combining qualitative and quantitative methods. Mixing methods, then, necessarily entails mixing thick and thin conceptualizations. But concepts cannot be translated from one space to another without significantly altering their dimensions. Identifying a case as “less” or “more” than another on an indicator scale is essentially different from depicting it as belonging to a particular ontological category. Creswell and Clark state that “unquestionably, it is easier to transform qualitative data into numeric counts (quantitative data) than vice versa. Transforming qualitative data involves reducing themes or codes to numeric information, such as dichotomous categories” (Creswell & Clark 2006: 128). But treating categorical concepts as dummy variables—a specific kind of scalar unit whose values is set either “1” if an attribute is present and “0” if it is absent—represents a significant sacrifice in analytical power because they only influence the constant term of a regression equation, not the slope. Moreover, such categorical variables cannot be arrayed in a meaningful ratio format, reducing their explanatory power in regression analysis (Taagepera 2008: 58, 229-31). The use of fuzzy-sets instead of dichotomous categories reduces the discrepancy between the thickness of qualitative and and thinness of quantitative conceptualization, but still relies on numeration (Verkuilen 2005). Ultimately, as Franzosi writes, “to measure is fundamentally to think relationally… different measurements of an object will lead to different relations of that object to other objects” (Franzosi 2004: 281).

Stretching between qualitative and quantitative methods manifests itself as a mismatch in the intension and extension in qualitative and quantitative concepts. While concepts may share a label or term, they have different secondary level characteristics
and refer to different categories of cases. This is shown in diagram 3. On one hand, the overlay of thick and thin concepts leaves numerous cases that would be counted as “in” in a quantitative setting but “out” in a qualitative setting. On the other hand, characteristics that are crucial in the qualitative definition of a concept are absent in the quantitative one.

Diagram 3

This problem is rarely acknowledged, much less addressed, even in the most astute works of MMR. Consider Lieberman’s study of the impact of race and regionalism on tax regimes (Lieberman 2001). Lieberman contends that different ideas of national political community (NPC) make the certain racial, religious, ethnic, or regional identities more likely to become politically salient than others:
The specification of group rights in the form of official state documents and policies provides a stronger set of incentives for political entrepreneurs to make claims based on such identities… Federalism, for example, tends to give important political salience to regional identities, and official racial exclusion tends to give much more salience to racial identities… (Lieberman 2001: 14).

These different forms of identity have an impact on the ability of political entrepreneurs to demand the creation of progressive or regressive tax systems.

Using a periodized, paired comparison of Brazil and South Africa, two states with high racial and regional disparities but different legal definitions of NPC, Lieberman traces a counterintuitive but persuasive historical logic: In Brazil, the 1891 Constitution adopted a federalist system that privileged claims based on regional equity, but was explicitly inclusive on racial grounds. Brazil’s white economic elite treated the state as an adversary and worked to block demands for greater redistribution. In contrast, South Africa’s 1909 Constitution specified white racial supremacy while denying recognition to regional differences. By establishing whites as a formally recognized legal category, this cornerstone legal document encouraged the white economic elite to cooperate with the state in establishing a social safety-net system that raised the living standards of their poorer co-ethnics and increased the solidarity of the ruling white minority. After apartheid’s downfall, however, this redistributive regime was opened to all races, turning a tool of racial exclusion into one of socio-economic amelioration.

Following the case studies, Lieberman deploys large-N regression analyses to test whether the similar legal definitions of NPC along racial and regional dimensions have
the same effect in other cases. Examining constitutions and other legal documents from over one hundred cases, he codes the data, converts them into a series of dummy variables, and enters them on the left hand side of the equation. The statistical results show a strong correlation consistent with the small-N study. When constitutions and other founding documents enshrine federalism, states tend to have less extraction capacity; when they enshrine racial supremacy, states have higher extractive capacity.

But NPC is stretched severely in this effort to shift from a thick to thin conceptualization. Lieberman initially focuses on the construction of political communities “in response to pre-existing racial and regional cleavages: whether or not to adopt federalism, and whether or not to continue excluding blacks from full and equal citizenship” (Lieberman 2001: 14). The thick conceptualization of NPC as an independent variable allows Lieberman to make a specific hypothesis about the basis for collective demands for economic redistribution on the part of political entrepreneurs. But this qualitatively-derived elaboration of NPC proves too narrow to incorporate the majority of empirically relevant cases. As shown in Tables 2 and 3, South Africa and Brazil are only representative of six of the sixty-nine (8.5%) of cases included in the quantitative population. Only twenty-one countries (30.4%) in the quantitative population share South Africa and Brazil’s racial fragmentation. Forty-three cases in the population (62.3%) have neither relevant racial or regional cleavage, meaning they essentially are outside the initial definition of NPC used in the qualitative narrative. In the quantitative analysis, Lieberman incorporates these cases by inventing a new residual category (“non-fragmented”) for cases where racial divisions are not present in either formal or informal form, but the overall explanatory power in the study proves
profoundly uneven. Based on Goertz’s possibility principle, the forty-three cases that do not have pre-existing racial or regional cleavages should be excluded from consideration as irrelevant, since they do not share the characteristics necessary to trigger the posited mechanism (Goertz 2006). The cases of federal-exclusionary and unitary-inclusionary are similarly problematic; since so much of Lieberman’s qualitative argument revolves around the interaction of these two variables, explaining their independent impacts is difficult. Ultimately, where Lieberman’s account of the mechanisms involves in building tax regimes in countries with a combination of ethnic diversity and regional division, it is essentially mute regarding the majority of cases in which racial or regional cleavages are unavailable for political entrepreneurs to exploit.
Table 2: Initial Typology of National Political Community and Distribution of Cases by Type in Qualitative Analysis from Lieberman (2001: 79)

<table>
<thead>
<tr>
<th>Type</th>
<th>Federal</th>
<th>Unitary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race exclusionary</td>
<td>South Africa</td>
<td></td>
</tr>
<tr>
<td>Race inclusionary</td>
<td>Brazil</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Distribution of Cases in Quantitative Analysis from Lieberman (2001: 242)

<table>
<thead>
<tr>
<th>Fragmented</th>
<th>Race exclusionary</th>
<th>Federal</th>
<th>Unitary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Race inclusionary</td>
<td>2 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Fragmented</td>
<td>7 43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similar problems of conceptual stretching are evident in Schultz’s study of democratic peace, where statistical observation of correlational regularities precedes qualitative investigation of mechanisms (Schultz 2001). Schultz uses POLITY scores to define a domain of some fifty-eight cases in which democracies faced challenges of deterrence and codes whether or not the opposition stood with the government during the crisis and whether the deterrence, was a success. Through chi-squared and probit analyses, he demonstrates that success or failure is closely correlated with whether or not the democratic opposition sides with the government during a crisis.

Based on these statistical findings, Schultz examines the British in the Fashoda Crisis, the British in the Boer War, the British and French in the Rhineland Crisis, the British and French in the Suez Crisis, and the British during Rhodesian Independence to trace the way opposition behavior tips the government’s hand and reveals information to foreign rivals about the level of determination in the crisis. He concludes that whereas nondemocratic governments have substantial leeway to bluff and probe, democratic governments are less willing to make threats they do not intend to carry out. Underlying this probabilistic prediction is a
specific causal mechanism: democratic governments face domestic competitors who have an incentive to oppose the use of force when political and military conditions are unfavorable (Schultz 2001: 233).

When the opposition supports the government it effectively lends credence to the national leadership’s claims of resolve in a crisis and chastens potential challengers. When the opposition stands apart from the government it damaged the government’s credibility.

Nearly all of the inferential leverage of the qualitative portion of the study, however, comes from examining a single political system, Britain’s Westminster government, a unique construction in terms of transparency of opposition and autonomy of the government. As Peterson (1996) and Elman (1997) note, this system has differences between presidential and parliamentary systems, in the institutions of civil-military interaction, and the specific constellation and strength of opposition forces at the time of crisis, all impinge on the ability of a democratic opposition to check the government’s autonomy in formulating foreign policy. The diversity in democratic institutional design renders the hypothesized signaling mechanism improbably. In the United States, for instance, the concentration of foreign policy decision-making in the relatively opaque and autonomous executive branch excludes the opposition from foreign policy debates and renders moot most signals that emanate from the legislative opposition. Just as for Lieberman, the coincidence between statistical coefficients and qualitative findings is made less persuasive because of the apparent heterogeneity of the cases about which he seeks to generalize.

Different research objectives dictate different sequencing in the use of qualitative and quantitative techniques (Creswell & Clark 2006). In explanatory studies
like Schultz’s, the key concept of democracy and contentions of democratic peace theory were already well-known, so his project bypasses serious conceptual development and moves directly to demonstrating a statistical correlation and then uses qualitative methods to explain it. In exploratory studies like Lieberman, the independent variable (i.e., NPC) is itself a novel concept that has to be explored qualitatively before it could be tested again in a quantitative format. In both of these examples, though, the initial choice of qualitative or quantitative conceptualization binds the researcher to a particular conceptualization which has to be alternatively thickened or thinned. This erodes the confidence of MMR findings. Conceptual stretching must be addressed more directly to make MMR a viable research strategy.

Making MMR Work

Can MMR proceed with inconsistent conceptualization of variables? Does conceptual stretching doom MMR? The implications of conceptual stretching depend on the logic upon which MMR is premised. Among the earliest justifications for mixing qualitative and quantitative empirical studies is the logic of validation through triangulation of results (Ragin 1989; Denzin 1978). Put simply, two tests are better than one, since “a hypothesis that had survived a series of tests with different methods would be regarded as more valid than a hypothesis tested only with the help of a single method” (Erzberger & Keele 2003: 460). An exemplar of this logic comes from a recent report on combining quantitative and qualitative models to predict state failure, in which Goldstone relates that “having at least two independent approaches to assessing instability, if they
point in the same direction, greatly increases the confidence of predictions” (Goldstone 2008: 7).

There have been some methodological critiques of this proposition. Qualitative and quantitative findings may have compounding rather than correcting biases and thereby grant no warranty of validation (Fielding & Fielding 1986; Rohlfing 2007). Conceptual stretching, however, is a more fundamental and fatal flaw. If what is categorized as “instability” (or democracy, or any other variable) in a qualitative setting is not equivalent to what is scored as such in a quantitative setting, congruence between the findings of a qualitative and quantitative study can be dismissed as merely a felicitious coincidence. The only solution to this problem is definitional transitivity, in which the conceptual definitions are explicitly stated to be synonymous and case domains equivalent. But, as demonstrated by Schultz, this quickly degenerates into a situation in which a country is deemed a democracy simply because it meets whatever numerical cut-point POLITY or Freedom House datasets deemed necessary to be grouped as such, not because of any innate relationship between the case and the concept of democracy itself. Such a delineation of a universe of cases comes at the cost of stripping the underlying concept of its connotative specificity.

Another approach to MMR, however, exploits the inherent gap between qualitative and quantitatively-derived concepts to produce complementary—not mutually validating—causal accounts. As Mahoney and Goertz relate,

an explanation of an outcome in one or a small number of cases leads one to wonder if the same factors are at work when a broader understanding of scope is adopted, stimulating a larger-N analysis in which the goal is to
explain particular cases and more to estimate average effects. Likewise, when the statistical results about the effects of causes are reported, it seems natural to ask if these results make sense in terms of the history of individual cases (231).

Gerring notes that “all causal arguments strive for evidence of covariational (correlational) relationships between the putative $X$ and $Y$ as well as evidence of causal pathways between $X$ and $Y$” (Gerring 2005: 191; see also George & Bennett 2005: Chapter 7).

The logic of complementarity in MMR is the ability to connect macro-level structural variables with micro-level mechanistic accounts of agent-based action. The goal is a more comprehensive causal account using quantitative techniques (and variables) to isolate regularities and qualitative techniques (and variables) to identify the processes that link antecedents and outcome.\(^2\) An example of this approach comes from Sambinas’ (2004) discussion of applying MMR to the study of civil war, which differs considerably from Goldstone’s search for mutual validation. Take, for instance, the impact of education on civil war: Statistical evidence shows that states with lower levels of education have a greater likelihood of suffering from civil war. The larger the pool of unemployed and poorly educated people in a country, the more people who stand to gain by seizing booty during war rather than other economic pursuits. But some countries with relatively well-school citizens, like Cyprus, Yugoslavia, Georgia, Russia, and Lebanon, all experienced civil war, while Saudi Arabia, with a very low education rate, has not. To make sense of this anomaly, Sambinas notes, we need to explain the manifest

---

ways schooling might influence civil war risk. Qualitative studies suggest that it is not just the quantitative extent of schooling, but the qualitative type of education, which can influence the propensity for war. When an educational curriculum disseminates ethnic chauvinism, for instance, it actually reinforces the motive for violence and increase the likelihood of civil war (Sambinas 2004: 264; see also, Davis 2005). Thus, the same variable can be associated with very different outcomes depending on whether the underlying concept is treated qualitatively or quantitatively.

Effective combination of qualitative and quantitative concepts is the lynchpin in complementary MMR, linking structural conditions and agent-based actions. But in so doing, MMR must address the problem of conceptual stretching head-on. The assumption that underlying populations are congruent and homogenous must be abandoned. Instead, the burden falls on the researcher to examine inductively the incongruence in intention and heterogeneity in extension. Initially, she may follow Dunning’s recommendation that a researcher code five or ten country-cases manually before using a pre-packaged dataset in order to ensure familiarity with variable features and underlying conceptual definitions (Dunning 2007). This is especially important when potentially disparate qualitative and quantitative versions of a concept are in play.

MMR, however, also requires a more rigorous theoretical justification for the case selection. Various scholars have advocated random selection, focused paired comparisons, or critical cases as techniques to identify cases for intensive qualitative interrogation from amongst a quantitatively-defined population (Lieberman 2005). For the purpose of conceptual clarification, though, the most relevant cases are the most

---

3 See also the symposium on case selection, case studies, and causal inference in *Qualitative and Mixed Method Research* (Fall 2008).
extreme or liminal at the independent or dependent variables. It is the disposition of these cross-over cases and the determination about whether they can be considered democracies (or not), or to have experienced civil war (or not), that is the most crucial from the perspective of concept formation because these cases highlight precisely the criterion of inclusion and exclusion within a categorical set (Ragin 2000). Once again, the goal is not to amend qualitative and quantitative conceptual definitions until they are equivalent, but to sound-out their divergences.

Explicating empirical heterogeneity and conceptual incongruity is vital to capturing the multiple and manifest pathways connecting structural antecedents and outcomes. Like all studies using narrative techniques to trace causal mechanisms, MMR must establish the truth-claim of its narrative analysis and test alternative explanations. The more instances which demonstrate a single analytically analogous process from initial condition to outcome, the more compelling are the mechanistic arguments (Büthe 2002). But identifying a single process is rarely enough to account for the full range of correlation observed in the quantitative analysis. In the examples from Lieberman and Schultz cited above, the problem is not that the mechanisms specified by qualitative methods are incompatible with the statistical pattern. Rather, the problem is that the account is incomplete: based on the possibility principle, these mechanisms are logically precluded from functioning in many of the cases in the large-n population. Stated another way, a significant portion of the $R^2$ observed in the quantitative portion must have been due to mechanisms other than the ones identified in the study. Had Schultz recognized that the British Westminster-system was but one type of democratic regime within the population which possessed unique properties in terms of transparency, he
could have continued to look at other mechanisms to account for peaceful dispute resolution among other democratic regime forms. For Lieberman, once it was clear his initial conceptualization of NPC was too limited to accommodate all the observed cases, he could have conducted case studies in all six quadrants of the NPC typology in order to ascertain the mechanisms leading to variation in tax and redistributive regime in cases where extreme ethnic diversity was absent. More attention to the exact parameters of stretching between qualitative and quantitative concepts allows the researcher to look for these alternative mechanisms.

Since inferences about mechanisms are derived from the properties of the concept used in the qualitative analysis, the power of MMR is constrained by the representativeness of the qualitative sample. Case studies are not handmaidens to the quantitative model, simply used to confirm the directional findings of a statistical model or to identify omitted variables (Sil 2004). Rather, qualitative concepts and methods are the fulcrum, highlighting heterogeneity among case units and locating multiple mechanistic pathways connecting from initial conditions to outcomes (Teddlie & Yu 2007). The success of complementary MMR hinges on exploring the overlap between large and small-n conceptualization. Thick and thin concepts should not be made equivalent or merged by deductive theoretical reasoning. Rather, inductive techniques must be used to demonstrate the contours and limits of this analytic space.

Conclusion

Mistranslation between qualitative and quantitative components is inherent to any MMR project. Scoring and categorizing are not equivalent. The trade-off of intension
and extension leads ineluctably to conceptual stretching. Stretching poses a severe problem for MMR which attempts to gain validation through triangulation. Simply stating that the coefficients and p-values in a regression table match the predicted signs from a case-study is not saying much if the variables in question do not measure the same object or refer to the same cases.

MMR’s greatest promise comes in seeking explanations that identify both structural correlations and causal processes. Such accounts are not necessarily more valid, but are more comprehensive than those offered by a single method, covering both a wider and more intricate span of the causal pathway (Gerring 2005; Roberts 1996). By its very nature, MMR disavows attempts to merge quantitative and qualitative methods and unify structural and mechanistic components, such as is sought in narrative sequence analysis (Abbott 1995). In retaining the autonomy of case studies and statistical techniques, MMR must explicitly address the strengths and weaknesses of each methodological component, including the implications of conceptual stretching. Instead of seeking to make thick and thin conceptualizations equivalent, it should seek ways to use their respective differences to make inferences at both the macro- and micro-level.
Bibliography


