BUFFETT INSTITUTE
NORTHWESTERN UNIVERSITY

Buffett Institute Series Working Paper

Integrating Qualitative Causal Inferences into Policy-Oriented Development Research

Andre Nickow
PhD Student, Sociology, Northwestern University

Working Paper No. 15-001
June 2015

Andre Nickow

Integrating Qualitative Causal Inferences into Policy-Oriented Development Research

Abstract

In order to make informed policy recommendations, development research faces two broad challenges: first, to identify and estimate the effects of interventions where they have occurred, and second to build and test theories about the effects that a related intervention would be hypothesized to exert in a different context. I review three recent strands of research that seek to integrate qualitative methods into policy-oriented development research and find that qualitative methods, when incorporated into mainstream development research, are often used to improve measurement validity and to develop hypotheses and theories that address the first challenge. However, qualitative approaches to causal inference are rarely used, and as a result qualitative methods can rarely be used for testing theories, particularly with regard to the second challenge. After suggesting an a priori *argument for why qualitative causal inference could benefit policy-oriented research, I demonstrate how this approach gives rise to significantly different policy implications than would otherwise emerge in a specific context: rural land institutions in India.*

Bio

Andre is a Ph.D. student in Sociology at Northwestern, focusing on the political economy of development. He focuses on pro-poor growth, social movements, and civil society, with a regional concentration on South Asia.

Introduction

In order to make empirically grounded and theoretically informed policy recommendations, development research faces two broad challenges: first, to identify and estimate the effects of interventions where they have occurred, and second to build and test theories about the effects that a related intervention would be hypothesized to exert in a different context. A single paper, study design, or even research program need not, of course, take on both tasks. Indeed, the vast majority of research within development studies focuses on the first challenge, limiting itself to broad and provisional policy recommendations, if directly addressing the second challenge at all. But, as is often acknowledged, researchers seeking to evaluate an intervention in a way that will be helpful to future policy designers and other stakeholders will be more successful to the extent that their findings reveal not only the average effects of a black-boxed intervention on a sample, but also the mechanisms through which they occurred and the contextual features that govern when and where these mechanisms will operate outside the study sample. Implicitly or explicitly, within a single study or across a broad literature, policy-relevant development researchers (collectively) should strive both to accurately estimate effects of interventions and to explain specific, contextualized outcomes. That way, findings can be not simply "generalized", but also leveraged to make specific policy decisions with regard to other cases. At stake is the way that external validity is conceptualized and operationalized in the process of drawing policy inferences.

Each of these two challenges requires the careful and systematic use of empirical evidence to draw causal inferences, and each of these could conceivably be tackled with a wide variety of methodological tools to gather and organize data, as well as a variety of analytical lenses through which to consider causality. Although quantitative—econometric, experimental,

and quasi-experimental—methods have long predominated in their field, development scholars increasingly agree that researchers (at least collectively) should draw on a variety of the plethora of quantitative and qualitative methods and approaches that social science has to offer (Thomas 2008). As expressed by two World Bank researchers, "qualitative and quantitative approaches in development research and program evaluation can help yield insights that neither approach would produce on its own" (Rao and Woolcock 2003, 165). They seek to advance the "use of qualitative methods to generate more and better quantitative data and to understand the process by which an intervention works, in addition to ascertaining its overall impact" (169).

Within the past two decades especially, certain qualitative methods, including interviews, focus groups, ethnography, and historical case studies) have become more and more accepted by powerful development agencies, including the World Bank. This has given rise to a new wave of scholarship that productively engages research practices traditionally seen as on opposing ends of the qualitative-quantitative divide into dialogue with one another. Yet, qualitative and quantitative approaches in many ways continue to represent contrasting "cultures", with opposing norms, standards, and logics (Goertz and Mahoney 2012). When particular practices from across the divide are integrated, they may fail to speak coherently to one another. Or, alternatively, the resulting paradigms may tend to fall primarily within one camp or the other (Mahoney and Goertz 2006, 165-166). It is thus essential when reviewing mixed-methods development literatures to clarify the specific function each method plays in the overall research design, and in references to the two challenges of policy-oriented research described above.

In the present paper, I review some recent prominent attempts by development scholars to bring qualitative methods into their policy-oriented development research. I ask, what role do specific methods from each camp play in constituting the overall research and theory? By

policy-oriented development research, I mean any empirical social science research aimed at explaining variation in economic or social development that could hold implications for the practices of stakeholders, whether governments, agencies, or civic organizations. Although actual policy design involves normative judgments beyond the scope of social science, much development research strives to inform practitioners, whether directly or indirectly, implicitly or explicitly, of the most likely results of a particular intervention that will ultimately have to be implemented in a particular context.

My review reveals that qualitative methods, when incorporated into mainstream development research, are often used to improve measurement validity and to develop hypotheses and theories. However, qualitative approaches to causal inference are rarely used, and as a result qualitative methods can rarely be used for testing theories. The hypotheses developed in these mixed-methods studies and research programs tend to be formulated as probabilistic statements of average effect as is common in quantitative research rather than the deterministic statements of necessity and sufficiency that arise from qualitative research (Goertz and Starr 2003; Mahoney 2008). The central intent of this paper is to make the case that, despite the enormous productivity of existing mixed-methods research relative to exclusively quantitative designs, the neglect of qualitative approaches to causal inference is unfortunate.

In the remainder of the paper, I begin by reviewing recent mainstream critiques of the overreliance on quantitative methods in policy-oriented development research, in order to identify the problems that the methodological innovations discussed here attempt to solve. I then consider recent attempts to address these shortcomings using qualitative methods, and find that qualitative methods are often used to increase measurement validity or to improve theory development, but quantitative methods of causal inference continue to predominate in theory

formulation and testing. My review reveals that these tendencies hold for studies at both the micro and macro levels, i.e., for studies that focus on more immediate short term development programs, projects, and policies, as well as longer term historical studies that look at larger institutional changes. For the former category, I focus on the Q-Squared and participatory econometrics approaches, while for the latter I focus on the work of "historical economists" (see Woolcock, Szreter, and Rao 2011). Finally, I consider what it might look like to integrate qualitative causal inference into development research, and why this could contribute to better-informed policy. I conclude by illustrating the drastic reorientation in policy implications that may arise from this exercise in the context of a concrete example—land revenue institutions in India.

The Limitations of Quantitative Policy-Oriented Development Research

Since its inception in the mid-twentieth century, the field of development studies has been overwhelmingly dominated by the discipline of economics. While social scientists from all disciplines have dealt extensively with the substantive issues of development, economists have generated by far the largest body of work that places significant emphasis on policy and that policymakers have most listened to. It is no surprise, then, that the predominant tools of economics—econometrics, experiments, and quasi-experiments—have also predominated within the field of development studies. Economist Ravi Kanbur's (2002) observation that "Development economics stands in beleaguered ascendancy, atop development studies and development policy" (477) is only slightly less true today than when he wrote them.

This is not to say that the field's methodology has remained static or stagnant—on the contrary, consensuses have changed regularly with regard to particular models, specification

techniques, and standards. Most recently, stringent experimental and quasi-experimental models have come to the forefront. Quantitative development researchers are increasingly expected to follow or approximate experiments more closely than they were in the past. At the forefront of this tendency has been the recent explosion in field experiments (Banerjee and Duflo 2009; 2011; Humphreys and Weinstein 2009). To supplement field experiments, or to substitute for field experiments when they are impossible or impractical, development researchers also rely more heavily than ever on identification techniques that closely approximate experiments, including instrumental variable (IV) analysis (e.g., Narayan and Pritchett 1999; Qian 2008) and regression discontinuity analysis (e.g., Duflo, Dupas, and Kremer 2011). However, although these trends represent changing standards of statistical identification, the model of inference remains the same. Nearly all econometric techniques attempt to match the causal inference capabilities of experiments as closely as possible, isolating the average effects of a particular intervention across cases in a sample (Angrist and Pishcke 2001; 2009; Mahoney, Kimball, and Koivu 2009).

However, even as some development scholars pushed for ever greater commitment to experimental designs during the first decade of the twenty-first century, a less visible group (albeit one that is still more or less in the development mainstream), began to elaborate critiques of development research's overreliance on quantitative methods. Prominent among these have been several researchers who are or have been affiliated with the World Bank, including Ravi Kanbur, Vijayendra Rao, Martin Ravallion, Paul Shaffer, and Michael Woolcock. Why do these scholars argue that the predominant experimental methods are insufficient on their own to compellingly inform policy? In the remainder of the present section, I review their critiques of

exclusive reliance on quantitative methods, dividing them into critiques of measurement practices and critiques of theory development.

Measurement

A common critique of quantitative research methods is that they lack the ability to accurately measure certain complex or difficult to access phenomena. While some variables are straightforward and easily captured in regression equations with minimal loss of internal validity, others are not so easily captured, such as social capital and institutional strength. Even seemingly simple and straightforward rankings of opinions and perceptions are not necessarily straightforward to interpret. As Rao and Woolcock point out, "Many of the most important issues facing the poor—their identities, perceptions, and beliefs, for example—cannot be meaningfully reduced to numbers or adequately understood without reference to the immediate context in which they live" (2003, 167). Conceptual complexity has not stopped development researchers from attempting to study even the most complex constructs of social theory econometrically. They rely on a variety of ranking systems that attempt to indirectly capture the construct in question, for example, Knack and Keefer (1995) for institutions, and Knack and Keefer (1997) and Narayan and Pritchett (1999) for social capital.

A classic example of a concept that is difficult to measure quantitatively—common both within and beyond development studies—is democracy. Bowman, Lehoucq and Mahoney (2005) explain the threat of *"data-induced measurement error*", the "kind of error that occurs when analysts incorrectly code cases because of limitations in the underlying data on which they rely as description of empirical reality. Typically, [it] grows out of the use of inaccurate, partial, or misleading secondary sources."(940). Needless to say, the quality of an empirically based theory

can only be as good as the quality of the data used to build and test it. Debates about measurement error within poverty measurement have similarly raged within the field. In the edited volume *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*, nineteen development researchers, many of them employed by the World Bank, reflect on concerns of measurement error and measurement validity in poverty research, along with ways that qualitative methods can help to overcome at least some of these (Kanbur 2003).

Efficient survey design is difficult when the designer is far removed from on-the-ground substantive and communication dynamics (Rao 2002). Even aside from knowing which variables are most substantively relevant to the research concerns (a challenge that would fall more under theory development), it is difficult for survey designers to understand how respondents will most likely interpret questions, and thus the true meaning of the answers. For example, morbidity among the poor seems often to have been grossly underreported, since poorer respondents tend to have a lower standards for what constitutes good health (Sen 2002, 861; Shaffer 2013, 271). Similarly, a survey asking about domestic violence in India indicated that only 22% of women were beaten by their husbands, a number lower than comparable figures in the US. However, interviews revealed that the respondents interpreted beating to mean extremely severe violence resulting in serious injury. After refining the survey instrument to more accurately indicate the intended meaning of the question, rates went above 70% (Rao 2002, 1889).

In addition to its measurement validity being open to question, quantitative research on its own may suffer from more practical difficulties with surveys. For example, it is difficult to access certain populations with surveys (Rao and Woolcock 2003, 167), including slum residents (Jha, Rao, and Woolcock 2006) and people engaged in illegal behavior. Among these

populations, there is generally no way at the outset of a project to obtain an adequate sampling frame that even approaches the potential for establishing statistical representativeness. In these situations, qualitative methods like interviews and ethnography are needed as a substitute or as a point of access to approximating representative surveys.

Theory Development

Development scholars have also identified serious gaps in quantitative approaches with regard to certain elements of theory development. As noted above, experiments and the observational methods that seek to approximate them are designed to estimate average causal effects of a treatment within a sample on one or more dependent variable(s). Because of this, the experimental framework is limited in the extent to which it can contribute on its own to certain elements of theory development. By theory development, I mean the use of existing theoretical and/or empirical research to generate informed hypotheses for testing and interpretation. Shah and Corley (2006) capture the most common line of criticism in this vein as follows: "Theory building often requires the rich knowledge that only qualitative methods can provide…one of the most important limitations of cross-sectional, survey-based research is that it can only be used to test theory…Empirically grounded theory is most often developed through the use of qualitative methods as researchers generate a detailed understanding and thick descriptions of the phenomenon of interest; they collect information on many aspects of a phenomenon and attempt to document the perspectives of all key participants (1822). Quantitative researchers who shun qualitative work must still get their hypotheses from somewhere, so they tend to pull them from existing quantitative literatures or cursory qualitative case readings. As a result, as Rao (2002) suggests, "Ground-level realities can be ignored in favor of hypotheses that are constructed by

reading secondary literatures with relatively stagnant conceptualizations of human behavior…[this] aids in the reproduction of existing stereotypes rather than a constant updating of research questions based upon grounded understandings" (1887).

Theory development is a broad subject that can be understood from a variety of different perspectives. One element that is often important to the process is the inductive exploration of possible causal mechanisms. Close examination of particular cases can help to generate informed hypotheses that can then be tested through quantitative and/or qualitative methods. Identifying mechanisms involves attention to processes that unfold across space and time. I thus begin by looking at critiques of quantitative methods vis-à-vis their ability to contribute to theory development first with regard to context, the ways in which the circumstances of a place during a particular period of time may shape the interaction between variables of interest; and second, in terms of process, the way in which a series of events in a particular place evolve over time.

Context

First, the quantitative lens is limited in the extent to which it can contribute to developing theories about how the effects of a particular intervention may vary depending on context (Woolcock 2013). For Woolcock et al. (2011), this creates a self-defeating problem: "such forms of context-free policy sciences are severely handicapped as detailed guides to practical action in any particular context because the knowledge needed to speak to that context's specific local conditions and history…have been excluded by design from the policy model (Szreter et al., 2004: 12-13)" (79). Policy-oriented research that is exclusively quantitative tends implicitly to evaluate policies based on how they would function universally, rather than on how they would function in the presence of particular circumstances.

Development trends popular from the mid twentieth century through the 1980s, such as state-led development and then market-led development, infamously showed a disregard for context. They promoted "institutional monocropping," the adoption of homogenous organizational forms aimed at narrow targets of capital accumulation (Evans 2004; 2005). However, more recently it has become nearly axiomatic for economists to reject one-size-fits-all models (e.g., Banerjee and Duflo 2010; Easterly, 2014; Rodrik 2007) in favor of tailoring specialized solutions to diverse problems. For example, Sachs (2004) takes the metaphor of a physician, prescribing the right solutions for the right ailments, in his vision of "differential diagnosis". If, as most would now agree, context matters, it is a priority for development researchers to understand not only how interventions may work on average, but how they are likely to work differently in different contexts.

Process

Second, quantitative frameworks are alleged to be relatively weak at developing theories that can clearly specify temporal processes, including elements like timing and sequence. Rao and Woolcock write, "….quantitative methods…are less effective…in understanding process—that is, the mechanisms by which a particular intervention initiates a series of events that ultimately result in the observed impact—process issues—can be crucial to *understanding* impact, as opposed to simply *measuring* it" (2002,167). In an essay entitled "How and Why Does History Matter for Development Policy?", Woolcock et al. (2011) expand on this point extensively in the context of arguing for greater integration of history and related disciplines into policy-relevant development research, as do a variety of other scholars in the accompanying edited volume *History, Historians and Development Policy* (Bayly, Rao, Szreter and Woolcock

2011). Referencing the comparative historical literature in sociology and political science (Mahoney and Thelen 2009; Pierson 2004; Sewell 2005; Tilly 2002; important studies they do not cite include Aminzade 1992; Falleti 2005; Thelen 2000), these authors reiterate the point that temporal dynamics can take on their own autonomous causal efficacy in ways not easily captured by plugging a time variable into regressions, and advocate more efforts to confront these issues within the realm of policy-oriented development research.

Woolcock et al. (2011) furthermore point out that predominant econometric techniques assume "by default" that the "shape of the 'impact trajectory'" of an intervention is "monotonically increasing and linear", when this may be highly unlikely (80). Woolcock's (2009) essay is devoted to spelling out more specifically the dangers in such assumptions, along with prospective alternative approaches. Plotting two-dimensional graphs of time and impact, Woolcock draws on existing theory to come up with eight plausible trajectory shapes, but argues that the vast majority of econometric methods are efficacious only in capturing one of these.

With the critiques reviewed in this section as motivation, I next turn to considering how they are addressed through the integration of qualitative methods into policy-oriented development research.

Existing Approaches to Mixed Methods Development Research

How have development researchers critical of overreliance on econometric methods sought to address the methodological flaws? Individual methods across the methodological divide can conceivably be combined in a wide variety of ways, and researchers have taken numerous distinct approaches. Here, I focus on two broad mixed-methods approaches in mainstream development research: those that combine field methods with surveys or

experiments (the "Q-Squared" and participatory econometrics approaches) and an emerging tradition of econometric analyses that draw (at least to some extent) on qualitative historical analysis, practitioners of which have been usefully labeled by Woolcock et al. (2011) as "historical economists" (in order to distinguish them from the institutionalized sub-discipline of economic history).

Although many multi-method approaches exist, I focus on these sets of studies for two main reasons. First, these approaches have been spearheaded by mainstream, influential development researchers, as evidenced by the researchers' publishing regularly in top economics and development journals like *American Economic Review*, *World Development*, and *Journal of Development Studies* along with, in many cases, professional affiliation with the World Bank and other powerful development agencies. Second, they fall on different ends of the micro-macro spectrum. Q-Squared, participatory econometrics, and their cognate approaches tend to focus on shorter, more immediate periods surrounding development projects and policy changes, occurring more or less contemporaneously with the research. Historical econometric studies tend to take on longer time scales, and often focus on deeper and more gradual institutional change, in addition to or instead of clear-cut policies and programs. Together they account for a sizeable swath of contemporary mainstream development research that seeks to bring insights from qualitative methods into policy-oriented research. While reviewing each, I focus in particular on the way qualitative and quantitative methods are used at different functional points within the research design, including how they are used to address the two challenges of policy-oriented development research. This breakdown reveals that these approaches tend to incorporate qualitative data and methods for supplementary measurements and theory building, but rarely at the stage of causal inference. Given the centrality, significance, and

representativeness of the two strands of literature reviewed, this trend is likely generalizable to much mixed-methods development research currently taking place.

Qualitative Fieldwork in Mixed-Methods Development Research

Although select qualitative methods have been integrated into policy-oriented development research for several decades (see Lipton 1970 for an early example), mixed methods research designs within the field have seen explosive growth since around the year 2000 (Shaffer 2013a). One particularly prominent effort has been that of the Q-Squared initiative, a series of workshops and collaborations themed around mixed methods poverty research spearheaded by economist Ravi Kanbur, which has resulted in two monographs (Kanbur 2003; Shaffer 2013b), two journal special issues (Kanbur and Shaffer 2007; Shaffer, Kanbur, Thang and Bortei-Doku Aryeety, 2008), and numerous published journal articles. As suggested by the name, the goal is not simply to add a qualitative lens as an entirely separate (even if complementary) flow of information (presumably, such an effort would be called Q + Q). Instead, most of these researchers strive to go further, to integrate methods in a synergistic way so as to build a causal story of value that exceeds the sum of its methodological parts. Although researchers within this program have focused especially on a particular repertoire of strategies integrating quantitative and qualitative approaches to questions framed around or closely linked to "poverty", there are no set boundaries that sharply distinguish it. The present discussion therefore also applies to other studies following similar approaches that do not necessarily use the term Q-Squared.

In a recent, detailed review, Shaffer (2013a) divides the contributions of the first decade of the Q-Squared initiative into two broad areas: the "Identification Stage", which confronts the

questions of "who are the poor" and "what are their characteristics" (270), and the "Causal Stage", which attempts to explain "why [they are] poor (276). Shaffer's category of identification encompasses that of measurement (including measurement validity) discussed above in the present paper. Q-Squared efforts to transcend the weaknesses of large-N surveys by adding subjects' own perspectives on poverty into the study. Qualitative data is generally collected from a sub-sample through interviews, ethnography, focus groups, and especially participatory exercises in which, for example, groups of subjects may be asked to elaborate on the conditions of their basic needs and/or to rank households in terms of poverty. Many studies "standardize locally meaningful definitions [of poverty] and include them in a survey applied to a probabilistically sampled population" (272). Qualitative data can be used to design or redesign future survey instruments, or can be mapped onto other variables with a simultaneous survey. The qualitative data is used primarily for triangulating and adjusting the quantitative measures in order to improve validity.

The function of qualitative methods in the causal analysis of poverty for Q-Squared researchers is somewhat more complex. Shaffer categorizes the Q-Squared contribution in this arena into the following groups: "causal variables, weights, mechanisms, and the causal 'tree,' while also directing attention to external validity" (Shaffer 2013, 276). In other words qualitative analyses can generate insights on which factors exert causation, their relative importance, the sequence of causation between multiple variables, and the channel(s) through which they exert their causal effects. A prototypical approach (also common in other fields of mixed methods research) is to accompany a longitudinal survey with interviews, ethnographies, and/or case histories in order to identify plausible mechanisms that may explain the quantitative results. In other words, the idea is "to combine analyses of outcomes and processes" (276).

Shaffer cites the examples of studies by Place et al. (2007) and Woldehanna (2008), in which the authors use interviews and ethnographies to identify likely mechanisms of effect, explain counterintuitive findings, and substantively distinguish between outliers and non-outliers to explain patterns of agricultural technology adoption in Kenya and child labor in Ethiopia, respectively.

Within Q-Squared causal analysis, how, more specifically, are the two methodologies relating to one another? On one level, they can be seen as two separate statements, parallel causal stories. The quantitative side tests the average effects of variables (the outcomes) while the qualitative methods describe how they occur (processes). But to the extent that they do speak to one another, the qualitative data essentially function as a tool to triangulate or cross-check the quantitative findings. For the purposes of theory testing and external validity (which implies the policy question of how interventions might best be designed in other contexts), the quantitative analysis remains dominant within these study designs. Nonetheless, qualitative methods play a powerful role in shaping study findings.

A second line of research that overlaps significantly with Q-Squared has been the approach that Vijayendra Rao and collaborators call "participatory econometrics" (1998; 2002; Jha, Rao, and Woolcock 2006). As in Q-Squared research, these studies generally combine interview-based field methods with survey-based quantitative methods. For example, Rao and Ibanez (2005) integrate community case studies based on interviews and focus groups with a random sample survey to estimate and explain the effects of the Jamaica Social Investment Fund (JSIF), a government-run, community-driven development project sponsored by the World Bank. The authors identified five communities in which the fund operated, and used propensity score matching to identify a comparison community for each. They (along with a research team) then

carried out in-depth interviews and focus groups with community leaders and program participants from each community. Simultaneously, they administered a survey to fifty randomly selected households within each community. The qualitative data is used "to sketch a narrative of participation and collective action in JSIF and non-JSIF communities" (779), to observe mechanisms through which the program seemed to be exerting (or not exerting) its effects. The authors are careful to point out, however, that qualitative subjects were selected based on their roles within the community, and were not selected using statistical techniques to ensure representativeness. Thus, they rely on regressions based on the survey data to "understand the extent to which…qualitative findings can be generalized to the population" (810).

Although these next studies do not evaluate specific policies or development projects, Bloch and Rao (2002) and Jha, Rao, and Woolcock (2006) use similar research designs to generate policy-relevant findings about governance structures within Delhi slums and links between dowry disputes and wife abuse among a particular sub-caste in the Indian state of Karnataka, respectively. These studies, however, begin with in-depth, semi-structured interviews and ethnographies within their respective communities, and then use the qualitative data to design survey instruments and to develop theoretical models for testing with these instruments. For example, Jha et al. found early on that *pradhans* (informal leaders) played important roles in slum governance, but discovered only through the in-depth ethnographies that the roles these *pradhans* played tended to differ depending on whether the slum was an ethnically homogenous slum that had imported village governance structures, or a heterogenous slum in which relationships of governance were being built for the first time. This finding was then reflected in their survey (236). Relatedly, Bloch and Rao used qualitative data to shape (or

at least influence) their model specification, beginning their study by collecting ethnographic data and using this data "to inform the development of a noncooperative model of bargaining and signaling" (1029).

Given the recent pre-eminence of field experiments in policy-oriented development research, it is surprising that very few attempts within this field have used RCTs within mixed-methods research designs. One recent exception is a World Bank policy research working paper by Ananthpur, Malik, and Rao (2014), who combine a field experiment with ethnography to estimate and explain the results of a project intended to educate and mobilize villagers to more effectively and productively take advantage of a huge influx of government resources granted to *gram panchayats* (village councils) to local level development projects. Since the quantitative difference-in-difference comparison of treatment and control groups finds no statistically significant effects of the project, the role of the ethnography is to, as the title suggests, sketch "the anatomy of failure", to identify the mechanisms that most likely led to the overall null results. The analysis yields three explanations of failure that the authors argue are likely to have been responsible for the overall results: quality of facilitators (the primary administrators of the project); "lack of top-down support", and "persistent inequality" (i.e., legacies of feudalism).

Like Q-Squared, participatory econometrics studies tend to rely heavily on qualitative methods to increase measurement validity and inform theory development, but defer to quantitative methods when testing theories and considering external validity. Rao makes this point explicitly in the context of summarizing the contributions of participatory econometrics research designs: *"Qualitative data provide a context and aid in the interpretation of quantitative findings, while quantitative data check the generalizability of those findings"* (2002: 1891, emphasis in original). In terms of the two challenges for policy-oriented development

research described above, both Q-Squared and participatory econometrics approaches generally use qualitative methods only for the first.

Historical Economics

In addition to using qualitative interview- and observation-based field methods to supplement quantitative surveys and (occasionally) field experiments, economists (even outside the sub-discipline of economic history) have increasingly turned to qualitative historical data to supplement longitudinal econometric analyses. For example, even before the field experiment trend swept micro-oriented development studies, a strand of historically oriented econometrics-based research had begun a parallel wave in macro-oriented development research—often using quasi-experiments, especially analyzed through IV regression analysis, to study long term institutional transformations. Unlike the Q-Squared and participatory econometrics approaches described above, these studies do not usually conceptualize themselves as mixed-methods, and the role played by qualitative historical data may be small or large. Given the subject of this paper, I focus here primarily on studies in which the qualitative historical data is relatively prominent.

Most famous and influential among these historical economic studies have been the seminal papers by Acemoglu, Johnson, and Robinson (2001; 2002; 2005) on the long-term effects of colonial institutions, along with La Porta and collaborators' (1998; 2008; Djankov et al. 2003) work on the economic consequences of legal institutions. These lines of research emerged from related studies within the sub-discipline of economic history (e.g., Engerman and Sokoloff 2002; Sokoloff and Engerman 2000), but went beyond them in adapting more sophisticated techniques to approximate experimental models, thereby (ostensibly) bringing

historical orientation into the development economics mainstream. Although these early research programs were highly general in orientation and not particularly policy-oriented, they have inspired a new wave of studies, many of which are more specifically targeted to particular institutions or countries.

More broadly, history provides a wealth of cases that it would be unfeasible or unethical for field experiments to replicate. For example, forced migration as a result of violence, persecution, or environmental change is a common problem in the developing world, yet migration would not be ethical to induce. Bhardwaj and Fenske (2012) address this problem using historical data from the eastern Indian states of West Bengal, Bihar, and Orissa in the aftermath of Partition, as they faced a net inflow of around three million refugees from what was then East Pakistan (today's Bangladesh). Partition as a whole was undoubtedly a catastrophic event that caused enormous economic devastation, not to mention the human cost. But, Bhardwaj and Fenske ask, were there also economic benefits of the ensuing migration? Relying on geographical location indicators as instruments, their regressions show that higher numbers of East Pakistani immigrants in a given area were associated with economic boosts, most likely because they brought skills for growing jute. The authors take an uncharacteristically qualitative tone in their (albeit brief) discussion of scope conditions, proposing the need to identify necessary "opportunities", in this case met by available land and markets for cash crops (1086).

Another common use of qualitative historical data in these studies is the identification of instruments. Rodrik (2004) even suggests that Acemoglu et al. (2001)'s great success was not in coming up with or properly testing a compelling theory, but in proposing a clever and innovative instrument (settler mortality rates). The data for their instrument was provided by a historian, Philip D. Curtin, and could not have been discovered without the authors having read (or at least

been told about) the historiographic literature. Beyond the identification of the instrument, however, history takes a back seat to econometrics in the analysis. Similarly, but in a more focused and policy-relevant context, Duflo (2001, 2004) estimates the effects of education on wages in Indonesia using a natural historical experiment—timing differences in school construction across provinces during a massive government-led school construction project carried out between 1973 and 1978. Duflo is not particularly interested in the history of institutional development in Indonesia; rather, history provides a convenient tool. Presumably, without restrictions of cost and ethics, Duflo would have preferred to have orchestrated a massive field experiment. But, even if such an experiment were possible, this natural experiment is no doubt immensely more cost-effective. Nonetheless, this historical event constituted a unique methodological opportunity that could not have been discovered without significant consideration of historiographic literature. In addition to identifying instruments, historical analysis can help confirm the validity of an instrument, as shown in Banerjee and Iyer's (2005) study of land institutions in India—this study will be discussed in great depth below. In a book chapter for an edited volume based on this study (2010), they spend eight of 28 pages mining the Indian historiographic literature, in large part to make the case for the substantive validity of their instrument.

Another way in which qualitative historical insights can aid in historical econometric studies is by guiding theory development, for example helping to identify useful and empirically grounded hypotheses for testing. Kochar (2008) uses this strategy in attempting to explain the apparent failure of decentralization to improve the effectiveness of anti-poverty programs in India. While widely accepted explanations emphasize the relative political strength of the poor vis-à-vis elites (Bardhan and Mookherjee 2000; Dreze and Sen 1995), Kochar suspects that the

incentives provided to elites for pro-poor assistance may also be an important explanation for variation in success. To understand how this may occur, she turns to historiographic evidence of the British Old Poor Laws, which historians posit was a success in part as a result of the positive externalities of pro-poor assistance to the rich through labor market supply. This guides Kochar's econometric analysis, in which she uses the value of sugar subsidy per household as an instrument with which to test the effects of poverty assistance on labor market decisions, ultimately connecting labor market changes to elites' willingness to effectively allocate social welfare funds to poor households.

Qualitative Causal Inference: An Underused Tool for Policy-Oriented Development Theory?

As demonstrated in the above review, mainstream development researchers have attempted to integrate qualitative methods into their work in a variety of ways. Almost none of these studies, however, takes qualitative approaches to causal inference. Is this division of labor as it should be? Why might it be possible and productive to integrate qualitative causal inference into policy-oriented development research?

Economics as a discipline has virtually no accepted standards of qualitative methodology to speak of. However, several related disciplines have rich qualitative methodological traditions. Of these, perhaps the closest field to development economics is the area of political economy of development with disciplinary roots in comparative politics and sociology. Qualitative methods used by researchers within this field tend to center on case studies, whether contemporary case studies based on interviews, ethnography, and/or analysis of texts, or historical case studies based on primary and secondary documents (Bennett and Elman 2006; George and Bennett

2005; Woolcock 2013). Like econometricians, these qualitative methodologists often seek to draw causal inferences (Mahoney 2000). However, a key difference between the two approaches is that qualitative researchers tend to explain the causes of particular cases, while quantitative researchers tend to estimate the average effects of particular independent variables—in many cases a treatment or intervention—on a dependent variable of interest (Goertz and Mahoney 2012; Mahoney 2008).

What are the implications of this distinction for policy-oriented research? As suggested above, empirically and theoretically driven policy research must accomplish two tasks—first, to judge the effects of particular policies on outcomes of interest, and second, to use insights from these findings to draw inferences about what should be done in other contexts (no matter how cautious and provisional the second task may usually be in practice). Both of these tasks involve making causal inferences. But where do the types of causal inference the tasks respectively necessitate fall relative to the qualitative-quantitative divide as just described? Conventional methodological wisdom would hold that qualitative methods may play a role in measuring difficult to quantify outcomes, and inductively developing hypotheses about which factors might be most causally relevant. But it is often held as near axiomatic that quantitative tools of inference—and, in particular, statistical inference—hold sway in the realm of generalizability.

I suggest here that the optimal division of labor between qualitative and quantitative methods in policy-oriented research may be exactly the opposite: the first task fits more naturally with the quantitative logic of causal inference whereas the latter fits better with qualitative logic. When confronting the first challenge of policy research, one attempts to draw an inference about the average causal effects on an affected population, asking questions of the form: how did a given policy or program affect development? One could certainly draw on qualitative methods

in attempting to measure and theorize the effects of interest, and qualitative researchers do sometimes even address the "effects-of-causes" type questions as their central focus (Goertz and Mahoney 2012). But the overall causal logic nonetheless remains one of average effects.

Once the effect of an intervention has been inferred, one could simply go on to advocate those interventions found to have given rise to the most beneficial, most significant, and most cost effective results. This practice has been and remains common, and is in some ways implied by notions that quantitative causal inference should be the basis for generalization. But this strategy seems deeply flawed, if not outright absurd, in the context of the current consensus that one size does not fit all and that policymakers and stakeholders must tailor strategies to particular contexts. For a development practitioner confronted with a community and asked to recommend a policy, the ideal goal would not be to find the policy that had the best average effect on other communities (which might be more or less similar but certainly not identical). The goal would instead to identify the right policy for that particular community. As Deaton (2010) puts it, "the analysis of projects needs to be refocused toward the investigation of potentially generalizable mechanisms that explain why and in what contexts projects can be expected to work (426).

This identification process involves knowledge of mechanisms. If the mechanism through which an intervention led to an outcome can be specified, the researcher has an understanding of why it worked as it did, and thus a stronger basis for theorizing the circumstances under which it would or would not occur. While experimentalists often claim that experiments are the best, if not the only, way to establish the causality of a mechanism, the validity of this contention depends on the definition of mechanism used. Experimentally-oriented quantitative researchers usually understand a mechanism as an intervening variable that links a cause to an effect (Mahoney 2001, 578). Experimentalists can prove with a strong degree of certainty that a

treatment is a "mechanism" in this sense, since they randomly allocate the treatment. But the experiment will still say little about how the treatment exerted its effects, or which contextual factors facilitated the mechanism. I use the term mechanism to mean "an unobserved entity that—when activated—generates an outcome of interest" (581). Under this definition, mechanisms are unfolding processes rather than variables. This latter definition is more relevant to the challenge of policy-oriented development research described here—determining external validity and scope conditions in order to apply insights to highly diverse real world contexts.

The circumstances responsible for the presence (or absence) of the mechanism in question are generally conceptualized (whether implicitly or explicitly) as necessary and/or sufficient conditions (Goertz and Starr 2003), or aggregates of necessary and/or sufficient conditions, i.e., INUS and SUIN conditions (Mahoney et al. 2009). These conditions can be tested using any of wide variety of between-case comparative and within-case tools, including Mill's methods or variations of these (Mahoney 2007), process tracing (George and Bennett 2006; Mahoney 2012) and/or Qualitative Comparative Analysis (Ragin 1987; 2008). The researcher is thus better equipped to decide the best policy for a particular context, understanding the factors responsible for moderating the effects that will arise from the intervention in question.

A counterargument could be that, in a world of uncertainty and complexity, it is hard to find causes or causal combinations that always give rise to the same result. Instead, it may seem more realistic to interpret relations probabilistically. Indeed, quantitative researchers, once they have discovered an average effect across a sample, sometimes implicitly or explicitly interpret this average effect as a probability with which the factor will lead to a given outcome, an inherently illogical conception[1] that Mahoney (2008) refers to as the "definition of cause as

[1] A single case probability is said to be problematic because "…the real probability of the outcome is always equal to its ex post probability, which is 1 or 0" (Mahoney 2008, 416; see also Appleby 2004). This true probability (0 or

'probability raiser'" (415). Under this conception, one could feel safe in the assumption that an intervention will have a similar effect in one case as it did in other, similar cases. But the exact opposite could also be true, given the same set of quantitative findings. Mahoney uses the example of democracy in India to illustrate this point—it was the precisely the factors that make India statistically unlikely to be a democracy, such as ethnic fragmentation and inequality, that led its founders to put enormous effort into institutionalizing democracy, efforts that ultimately helped to explain the emergence of a stable democracy (416). The discussion below on land revenue institutions in India expand on this point extensively.

None of this is to say that the conventional methodological wisdom with regard to the division of labor between qualitative and quantitative causal inference is completely wrong or fallacious, but these *a priori* arguments should at least be a spur to consider more prospects for trying it the other way around. How might such an effort look in practice? In the following section, I reexamine Banerjee and Iyer's (2005) policy-oriented findings on land rights institutions in India and show how applying a qualitative approach to causal inference could lead to very different policy implications.

Land Rights and Class Conflict in India: A Lesson in Cooperation or Mobilization?

Banerjee and Iyer's (2005) *American Economic Review* article, "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India" elegantly blends (relatively) in-depth historical analysis with rigorous econometric techniques. The authors seek to contribute to the literature on the long-run effects of institutions on development (citing staple "historical economists" like Acemoglu et al., La Porta et al.). Unlike the former studies, they

---

1) can be guessed using a probabilistic fraction, but this fraction can only be derived from a sample with multiple (and probably many) cases.

narrow the focus to a particular institutional domain in a particular country, land revenue institutions in India. Although the time frame of their study stretches over a century, the study remains focused enough to generate insights of policy relevance.

During the colonial period, as the British gradually took over more and more of India, they implemented different land revenue institutions in different areas. There were three main models of land revenue institutions, but in the context of this study Banerjee and Duflo are interested primarily in the distinction between one of these—landlord-based revenue institutions (the infamous *zamindari* system)—contrasted against the two non-landlord revenue institutions (household-based or village-based systems). Under landlord-based revenue institutions, landlords were essentially given feudal property rights over the land they governed. The landlords were responsible for paying the colonial government a fixed amount, and beyond this were legally allowed to collect as much from tenants and agricultural laborers as they were able to extract. In contrast, under the non-landlord based systems, the state collected revenue directly from households and/or village representatives. Under all three systems, farmers were subjected to enormous revenue burdens, providing ample funding for the colonial state (Chandra et al. 2008). But landlord and non-landlord systems nonetheless represented qualitatively different institutions, as the former empowered a class of elite local intermediaries, while the latter two did not[2].

Banerjee and Iyer ask whether and how these different institutions had differentially affected the districts in which they existed by the 1980s, taking agricultural investment, agricultural yields, public investment (health and education) and health and education as dependent variables. Colonial land tenure systems were abolished immediately following

[2]Although some historical sources suggest that the village-based system could operate as a de facto landlord based system if a single family or oligarchy wielded enough power within a village.

independence in 1947, and all revenue intermediaries (i.e., the erstwhile *zamindars*) were in particular constitutionally prohibited. The power to implement land reform was transferred to individual states, and most of these passed significant amounts of land reform legislation (although the extent to which these land reforms were implemented in practice is thought to have been very limited). It is thus the historical legacies of these institutions, rather than the effects of the institutions themselves, being estimated in the study.

The study's central approach is, of course, econometric. The authors begin by establishing through OLS that former landlord districts significantly underperform former non-landlord districts on each of the dependent variables. To more compellingly demonstrate causality, they take two further steps toward approximating an experimental design. First, they estimate an IV regression model, using the fact of a district's having been conquered between 1820 and 1856 as an instrument for non-landlord status. Colonization during this time period is selected as an instrument because territories captured within this window were much more likely to be placed under non-landlord institutions, for a variety of historical reasons that the authors argue are unrelated to the dependent variable outcomes. The IV results confirm those of the OLS. Finally, the authors further confirm robustness by limiting their sample to territorially contiguous districts from which they construct matched pairs of landlord and non-landlord districts.

If the study were left here, it would not have much to offer in the way of policy implications, although these strong statistical findings would still be intriguing in light of several theoretical literatures within the political economy of development. It is of course far too late for long-dead British colonists to change their ways. But while the fact of an exogenous average (negative) effect of landlord institutions on long run development outcomes is all that the authors

infer with confidence, they do attempt to leverage their data to go even further and get a better sense of a more precise channel through which landlord institutions seem to have stunted development. Early in the paper, the authors present three prospective channels of influence. First, landlord institutions may have led to an increase in inequality, which, in turn stifled development. Second, landlord institutions may have created a conflictual political environment in which class conflict and mutual suspicion between landlords (or former landlords) and peasants, leading to poor development outcomes. Third, the colonial state may have invested more generously in infrastructure in non-landlord areas, since most landlord areas were governed by the Permanent Settlement, a law that made *zamindars* responsible for only a fixed rent. In contrast, the colonial state may have had greater incentives to invest in non-landlord areas, since they were more likely to be able to benefit from increased productivity.

Banerjee and Iyer conclude from the evidence that the second channel was the most likely one. While landlord-based areas did have much higher inequality levels immediately following independence, former landlord areas also underwent a larger number of post-independence land reforms, leading to a substantial (although not complete) convergence in inequality (but no accompanying rise in development for landlord districts). Thus, the high inequality engendered under the *zamindari* system, if it played a role in the long run underdevelopment of these regions at all, played only an indirect one. Differential infrastructural investment by the colonial state is even less likely to have played a role, since the regions most responsible for the non-landlord advantage show very little overlap with the non-landlord areas that received the most infrastructural investment. This leaves the second suggested channel—legacies of politicized class-based antipathy between large landholders and small peasants, as the authors' preferred scenario..

More specifically, the authors argue that class conflict continued to be higher in former landlord than non-landlord districts, even after the convergence in inequality levels following independence. This difference in class conflict grew into a difference in development outcomes following the 1960s. The mid to late 1960s and early 1970s saw a wave of enormous government programs aimed at rural development, including Intensive Rural Development Program, initiatives to spread new agricultural technologies and high yielding seeds that resulted in the Green Revolution, and India Gandhi's *Garibi Hatao* (End Poverty) campaign (1207-1208). Banerjee and Iyer argue that, as these national-level campaigns were going on, states consisting mostly of non-landlord districts were able to build the cross-class alliances necessary to take advantage of resources being offered. States composed mostly of former landlord districts on the other hand, according to this account, were too wrapped up in class struggles over land reform and other aspects of redistribution to achieve the collective action necessary to take advantage of central government resources.

The authors support this narrative with some further quantitative evidence. First, descriptive statistics reveal that the divergence did in fact occur primarily in the aftermath of the period in which these programs were implemented. Second, adding public spending into various regression specifications substantially reduces the non-landlord effect, indicating importance as a mediating variable. Having identified a concrete channel through which they are prepared to argue that landlord-based institutions led to poor development outcomes, the authors are in a position to draw a relatively direct policy inference: "the masses could perhaps have done a little better, or at least no worse, by focusing on what they had in common with the elites" (1210). Although it is too late to change political alliances during the decades of the 1960s-1980s, it requires no great leap to extend the policy question to one more relevant for today. Despite the

wave of liberalization and marketization that swept India during the 1990s, the central government continues to rely heavily on large public investment projects given to states as a central rural development strategy. Should NGO workers, local policymakers, and other stakeholders striving for rural development place more emphasis on empowering and redistributing resources toward tenants and smaller farmers? Or should they instead prioritize building cross-class collaboration in order to maximize public investment? Banerjee and Iyer's analysis clearly suggests the latter.

The methodology and logic of this study, specifically with reference to the ways it integrates qualitative approaches into the analysis, closely match the patterns discovered in the literature review above for development research as a whole. Qualitative historical analysis is used far more heavily in this study than it is in the vast majority of economic studies, including the work of the "historical economists", even including analysis of some primary documents (testimonies of British colonists who decided which land revenue institutions to implement). Qualitative historical analysis plays a central role in the study's theory development in several ways.

Most directly, the authors' use of qualitative historical analysis is essential for demonstrating that the fact of having been colonized by the British during the 1820-1856 is a valid instrument. To be a valid instrument and thereby establish causality according to quantitative standards, the instrument must be correlated with the dependent variable only through the independent variable, and the instrument must not itself be plausibly affected by the dependent variable. The authors use narrative analysis of primary and secondary documents to establish that decisions regarding which land revenue institution to implant were made not based on anything related to the land itself or its inhabitants. Instead, the decisions had to do with

ideologies predominant among (certain) British colonists, and the presence of precedents for non-landlord systems. For the most part, the British preferred non-landlord institutions after 1820 because it allowed for greater revenue, and because non-landlord systems were more feasible after this point since there were precedents and since the British had greater administrative capacity (non-landlord systems are more demanding of administrative capacity since the British had to collect from more parties). After 1856, there were few new territories left to be conquered, but the authors end the instrument period here since the British did switch the region of Oudh (within today's Uttar Pradesh province) back from non-landlord to landlord institutions as a favor to Oudh's leadership, which supported the British in putting down the Mutiny of 1857.

Despite the essential role of qualitative historical analysis within the study, the method of causal inference is fully quantitative. The authors seek to isolate the average effect of a treatment—non-landlord revenue institutions during the colonial period—on a population—districts in India colonized by the British. There is no attempt to consider a qualitative approach to causal inference, which in this case could mean determining how conditions relating to the colonial institutional legacies figured into broader chains of causal combinations (Mahoney et al. 2009). The study is fully convincing that, on average across the population, landlord-based colonial institutions had a negative effect on agricultural productivity and social development. But the channel through which this occurred, which is absolutely necessary to understand in order to infer valid policy implications—is less convincing. The authors are fully aware that the second step of their argument is less conclusive, and they attempt to put this in perspective early on in the paper: "our results would be more interesting if we could identify the reasons for this extreme persistence. While our data do not allow us to identify exactly the channel through

which the historical land revenue system continues to have an effect, there are a number of clues…." (1191). But readers interested in policy implications will likely remember these disclaimers much less than the implications that the authors are bold enough to infer.

How might qualitative causal inference change the policy implications generated by the study? As argued above, the second challenge of policy-oriented research is to draw causal inferences regarding what is most likely to work in a particular case. If the strength of qualitative causal inference is to explain particular cases, it would seem *a priori* appropriate to meet this challenge. By leaving readers, in the final sentence of their article, with the suggestion that class collaboration rather than redistribution would most likely have improved the development situation of former landlord areas, Banerjee and Iyer implicitly assume that a single intervention—cross-class collaboration—would have the same direction of impact. This assumption is in clear violation of the consensus in development studies against one-size-fits-all strategies. Adopting a qualitative causal inference approach would involve instead taking one or a few cases and attempting to fully explain them by uncovering the factors necessary and/or sufficient (or INUS or SUIN) for the desired mechanism to occur.

In the context of Indian land tenure institutions, an appropriate case unit of analysis would be the state, the level of government with the greatest constitutional authority to design and implement land policy. For states composed of mixed landlord and non-landlord districts, it would be difficult to clearly map findings onto the distinctions between institutional legacies introduced by Banerjee and Iyer. However, some contemporary states cover areas that are entirely or almost entirely composed of former landlord districts, and vice versa. India's third and fourth (respectively) most populous states—Bihar and West Bengal—cover the territory that once made up the *zamindari* heartland. But although they share the same score on Banerjee and

Iyer's key independent variable—the institutional legacy of *zamindari*—their outcomes have differed significantly. What insights can a comparison and explanation of their difference bring?

Although qualitative causal inference focuses on identifying the causes of outcomes in particular places, qualitative researchers still frequently make use of comparisons of two or a few cases, if not more. This is because some of their most powerful tools of causal inference are based on comparison (Goertz and Mahoney 2012; Mahoney 2007). Using this approach, we can seek to clarify the role of the mechanism proposed by Banerjee and Iyer—i.e., class conflict. If Banerjee and Iyer's study proves that institutional legacy is a cause and development an effect, then the mechanism which they suggest (but do not prove) connects the former to the latter is class conflict. But to say that, across all cases, more class conflict is associated with less development is not the same thing as saying that less class conflict would lead to improved development outcomes in a particular case. Yet Banerjee and Iyer's claim rests implicitly on the assumption that the two are more or less the same. Cross-case comparison using qualitative causal inference can check this assumption.

How do Bihar and West Bengal relate to one another with regard to the key variables of interest? During the 2007-2008 measurement year, West Bengal's Human Development Index (HDI) was at 0.492, above India's average of 0.467. Bihar's abysmal score of 0.367, on the other hand, placed it third from the bottom out of the 23 Indian states measured (India Human Development Report 2011, 25). Did West Bengal suffer less class conflict than Bihar? An initial review of the historical evidence suggests that it did not. If anything, West Bengal experienced more class conflict than Bihar, because in the former tenant and marginal formers formed strong class alliances, even across other communitarian divides (such as caste). These lower class alliances proved to be so strong, that they managed to force not only the legislation

but the implementation of massive tenancy reforms beginning in the late 1970s, a step that Banerjee himself documents elsewhere as strongly improving agricultural productivity in West Bengal (Banerjee et al. 2002). While by the 1970s West Bengal was facing a crisis of agricultural stagnation, tenancy reforms beginning in 1978—following an explosion of class-based mobilization that launched the Communist Part of India (Marxist) (CPM) into the state government—reversed this course and led to a marked improvement for the state's rural sector (Sengupta and Gazdar 1997). These reforms ended up being among the most successful of the dozens of reform efforts in India since independence (Sengupta 1981). Thus, far from being negatively associated with development in West Bengal, strong lower class mobilization in favor of redistribution seems to have been among the primary causes of the state's strongest success.

While Bihar also experienced class conflict in the sense of relatively poor factions competing for resources with relatively rich factions, non-class social divisions—and caste in particular—outweighed class, precluding a strong cross-caste lower class alliance as seen in West Bengal. Despite strong efforts and significant political will in favor of redistribution, as well as numerous land reform acts being passed, the lower classes were not able to force actual implementation of any land reform. In 1978, while Jyoti Basu's CPM government was implementing tenancy reforms in West Bengal, i.e., class-based redistribution, Karpoori Thakur's Janata Party government was implementing caste-based redistribution through a new caste reservation system in Bihar. This reservation no doubt channeled resources to marginalized groups that really needed them, but they also had the effect of splitting what once had been the lower class into two factions, further fractionalizing and disintegrating Bihar's lower classes (Blair 1980; Prasad 1979). In this case, caste conflict rather than class conflict seems to have led to Bihar's development failure. It is difficult to see how lower class coalitions,

which might have achieved the successful implementation of land reform, could possibly have made the situation any worse. To the extent that they may have overcome caste conflicts and challenged elite consolidation of power, it seems more likely that it would have helped.

Conclusion

While a great deal more historical data and rigorous analysis would be needed to fully establish alternative policy implications to those offered by Banerjee and Iyer (2005), the purpose of this discussion has been to show that qualitative approaches to causal inference, i.e., those revolving around the explanations of the mechanisms leading to outcomes in particular cases, may generate entirely different policy implications from approaches that seek to estimate the average effect of a particular intervention across a sample or population. Although I have presented some *a priori* arguments about why the qualitative approach might make more sense for the purposes of theorizing policy implications, the goal of this paper is not to put a stop to existing research practices, but to instead to encourage the deepening of a process that has already come along way within the past decade—the serious consideration of qualitative methods in policy-oriented development research.

Bibliography

Aminzade, Ronald. 1992. "Historical Sociology and Time." *Sociological Methods & Research*. 20(4): 456-480.

Ananthpur, Kripa, Kabir Malik, and Vijayendra Rao. 2014. *The Anatomy of Failure: An Ethnography of a Randomized Trial to Deepen Democracy in Rural India.* Policy Research Working Paper 6958. Development Research Group, Washington, DC: World Bank.

Angrist, Joshua D. and Jom-Steffen Pischke. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *The Journal of Economic Perspectives* 15(4): 69-85.

_____. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.

Appleby, D.M. 2004. "Probabilities Are Single-Case of Nothing." *Optics and Spectroscopy* 99: 447-456.

Banerjee, Abhijit V. and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics* 1(1): 151-178.

_____. 2011. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.* New York: PublicAffairs.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95(4): 1190-1213.

_____. 2010. "Colonial Land Tenure, Electoral Competition, and Public Goods in India." In Jared Diamond and James A. Robinson (Ed.s), *Natural Experiments of History*. Cambridge, MA: Harvard University Press, 185-221.

Bardhan, Pranab and Dilip Mookherjee. 2000. "Capture and Governance at Local and National Levels." *American Economic Review* 90(2): 135-139.

Bennett, Andrew and Colin Elman. 2006. "Recent Developments in Case Study Methods." *Annual Review of Political Science* 9: 455-476.

Blair, Harry W. 1980. "Rising Kulaks and Backward Classes in Bihar: Social Change in the Late 1970s." *Economic and Political Weekly* 15(2).

Bloch, Francis and Vijayendra Rao. 2002. "Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India." *American Economic Review* 92(4): 1029-1043.

Bowman, Kirk, Fabrice Lehoucq and James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38(8): 939-970.

Chandra, Bipan, Mridula Mukherjee and Aditya Mukherjee. 2008. *India Since Independence.* New York: Penguin Books.

Deaton, Angus. 2010. "Instruments, Randomization, and Learning about Development." *Journal of Economic Literature* 48(2): 424-455.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer. 2003. "Court." *The Quarterly Journal of Economics* 118(2): 453-517.

Dreze, Jean and Amartya Sen. 1995. *India: Economic Development and Social Opportunity*. Oxford: Clarendon Press.

Duflo, Ester, Pascaline Dupas, and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101(5): 1739-74.

Easterly, William. 2014. *The Tyranny of Experts: Economists, Dictators, and the Forgotten Rights of the Poor.* New York: Basic Books.

Engerman, Stanley L. and Kenneth L. Sokoloff. 2002. "Factor Endowments, Inequality, and Paths of Development Among New World Economies." *Economies*, 3(1): 41-109.

Evans, Peter. 2004. "Development as Institutional Change: The Pitfalls of Monocropping and the Potentials of Deliberation." *Studies in Comparative International Development* 38(4): 30-52.

_____. 2005. "The Challenges of the Institutional Turn: New Interdisciplinary Opportunities in Development Theory." In Victor Nee and Richard Swedberg (Ed.s) *The Economic Sociology of Capitalism*. Princeton, NJ: Princeton University Press, 90-116.

Falleti, Tulia G. 2005. "A Sequential Theory of Decentralization: Latin American Cases in Comparative Perspective." *American Political Science Review* 99(3): 327-346.

George, Alexander L. and Andrew Bennett. 2005. *Case Studies and Theory Development in the Social Sciences*. Boston, MA: MIT Press.

Goertz, Gary and James Mahoney. 2012. *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Goertz, Gary and Harvey Starr. 2002. *Necessary Conditions: Theory, Methodology, and Applications*. New York: Rowman & Littlefield.

Humphreys, Macartan and Jeremy M. Weinstein. 2009. "Field Experiments and the Political Economy of Development." *Annual Review of Political Science* 12: 367-378.

Jha, Saumitra, Vijayendra Rao and Michael Woolcock. 2006. "Governance in the Gullies: Democratic Responsiveness and Leadership in Delhi's Slums." *World Development* 35(2): 230-246.

Kanbur, Ravi. 2002. "Economics, Social Science and Development." *World Development* 30(3): 477-486.

_____. (Ed.). 2003. *Q-Squared: Qualitative and Quantitative Methods of Poverty Appraisal*. New York: Permanent Black.

Kanbur, Ravi and Paul Shaffer. 2007. "Epistemology, Normative Theory and Poverty Analysis: Implications for Q-Squared in Practice." *World Development* 35(2): 183-196.

Knack, Stephen and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures." *Economics & Politics* 7(3): 207-227.

Kochar, Anjini. 2008. "The Effectiveness of India's Anti-Poverty Programmes." *Journal of Development Studies*, 44(9): 1289-1308.

_____. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics*: 1251-1288.

La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins." *Journal of Economic Literature* 46(2): 430-465.

LaPorta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishyny. 1998. "Law and Finance." *Journal of Political Economy* 106: 1113-1155.

Lipton, Michael. 1970. "Interdisciplinary Studies in Less Developed Countries." *Journal of Development Studies*, 7(1): 5-18.

Mahoney, James. 2000. "Strategies of Causal Inference in Small-*N* Analysis." *Sociological Methods & Research* 28(4): 387-424.

_____. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16(3): 575-593.

_____. 2008. "Toward a Unified Theory of Causality." *Comparative Political Studies* 41(4/5): 412-436.

_____. 2012. "The Logic of Process Tracing Tests in the Social Sciences." *Sociological Methods & Research* 41(4): 570-597.

Mahoney, James and Gary Goertz. 2006. "A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research." *Political Analysis* 14(3): 227-249.

Mahoney, James, Erin Kimball, and Kendra L. Koivu. 2009. "The Logic of Historical Explanation in the Social Sciences." *Comparative Political Studies* 42: 114-146.

Mahoney, James and Kathleen Thelen. 2010. "A Theory of Gradual Institutional Change." In James Mahoney and Kathleen Thelen (Ed.s) *Explaining Institutional Change: Ambiguity, Agency and Power*. New York, Cambridge University Press: 1-37.

Narayan, Deepa and Lant Pritchett. 1999. "Cents and Sociability: Household Income and Social Capital in Rural Tanzania." *Economic Development and Cultural Change* 47(4): 871-897.

Pierson, Paul. 2004. *Politics in Time: History, Institutions, and Social Analysis*. Princeton: Princeton University Press.

Place, F., M. Adatao and P. Hebinck. 2007. "Understanding Rural Poverty and Investment in Agriculture: An Assessment of Integrated Qualitative and Quantitative Research in Western Kenya." *World Development* 35(2): 312-325.

Prasad, Pradhan H. 1979. "Caste and Class in Bihar." *Economic and Political Weekly* 14(7/8): 481-484.

Qian, N. 2008. "Missing Women and the Price of Tea in China: The Effect of Sex-Specific Earnings on Sex Imbalance." *Quarterly Journal of Economics* 123(3): .

Ragin, Charles. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University of California Press.

_____. 2008. *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago: University of Chicago press.

Rao, Vijarendra.1998. "Wife-Abuse, Its Causes and Its Impact on Intra-Household Resource Allocation in Rural Karnataka: A 'Participatory' Econometric Analysis.' In M. Krishnaraj , R. Sudarshan and A. Sharif (Ed.s), *Gender, Population, and Development*. Oxford: Oxford University Press.

_____. 2001. "Celebrations as Social Investments: Festival Expenditures, Unit Price Variation and Social Status in Rural India." *Journal of Development Studies*.

_____. 2002. "Experiments in 'Participatory Econometrics': Improving the Connection between Economic Analysis and the Real World." *Economic and Political Weekly* 37(20): 1887-1891.

Rao, Vijayendra and Ana Maria Ibanez. 2005. "The Social Impact of Social Funds in Jamaica: A 'Participatory Econometric' Analysis of Targeting, Collective Action, and Participation in Community-Driven Development." *Journal of Development Studies* 41(5): 788-838.

Rao, Vijayendra and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In F. Bourguignon and L.A. Pereira da Silva (Ed.s), *The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools*. Washington, DC, World Bank: 165-190.

Rodrik, Dani. 2007. *One Economics, Many Recipes: Globalization, Institutions, and Economic Growth*. Princeton, NJ: Princeton University Press.

Sachs, Jeffrey D. 2005. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin.

Sen, Amartya. 2002. "Health: Perception Versus Observation: Self-Reported Morbidity Has Severe Limitations and Can Be Extremely Misleading." *British Medical Journal* 324(7342): 860-61.

Sengupta, Sunil. 1981. "West Bengal Land Reforms and the Agrarian Scene." *Economic and Political Weekly* 16(25/6): A71-A75.

Sengupta, Sunil and Haris Gazdar. 1997. "Agrarian Politics and Rural Development in West Bengal." In Jean Dreze and Amartya Sen (Ed.s) *Indian Development: Selected Regional Perspectives*, New York: Oxford University Press, 129-204.

Sewell, William H., Jr. 2005. *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press.

Shaffer, Paul. 2013a. "Ten Years of 'Q-Squared': Implications for Understanding and Explaining Poverty." *World Development* 45: 269-285.

_____. 2013b. *Q-Squared: Combining Qualitative and Quantitative Approaches in Poverty Analysis.* Oxford: Oxford University Press.

Shaffer, P., Kanbur, R., Thang, N. and Bortei-Doku Aryeety, E. (Eds.). "Introduction to Q-Squared in Policy: The Use of Qualitative and Quantitative Methods of poverty Analysis in Decision-Making." *International Journal of Multiple Research Approaches* 2(2): 134-144.

Shah, Sonali K. and Kevin G. Corley. 2006. "Building Better Theory by Bridging the Quantitative-Qualitative Divide." *Journal of Management Studies* 43(8): 1821-1835.

Sokoloff, Kenneth L. and Stanly L. Engerman. 2000. "Institutions, Factor Endowments, and Paths of Development in the New World." *Journal of Economic Perspectives* 14(3): 217-232.

Thelen, Kathleen. 2000. "Timing and Temporality in the Analysis of Institutional Evolution and Change." *Studies in American Political Development* 14(1): 101-108.

Thomas, Bejoy K. 2008. "Methodological Debate in Poverty Studies: Towards 'Participatory Qual-Quant'?" *Development in Practice* 18(2). 280-288.

Tilly, Charles. 2002. *Stories, Identities, and Political Change.* New York: Rowman & Littlefield.

Woldehanna, T., N. Jones, and B. Tefera. 2008. "The Invisibility of Children's Paid and Unpaid Work: Implications for Ethiopia's National Poverty Reduction Policy." *Childhood* 15(2): 177-201.

Woolcock, Michael. 2013. "Using Case Studies to Explore the External Validity of 'Complex' Development Interventions." *Evaluations* 19: 229-248.

_____. 2009. "Toward a Plurality of Methods in Project Evaluation: A Contextualized Approach to Understanding Impact Trajectories and Efficacy." *Journal of Development Effectiveness* 1(1): 1-14.

Woolcock, Michael, Simon Szreter, and Vijayendra Rao. 2011. "How and Why Does History Matter for Development Policy?" *Journal of Development Studies* 47(1): 70-96.